UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE INFORMÁTICA

Web Interaction Environments: Characterising Web Accessibility at the Large

Rui Miguel do Nascimento Dias Lopes

DOUTORAMENTO EM INFORMÁTICA
ESPECIALIDADE ENGENHARIA INFORMÁTICA

2011

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE INFORMÁTICA

Web Interaction Environments: Characterising Web Accessibility at the Large

Rui Miguel do Nascimento Dias Lopes

Tese submetida para obtenção do grau de
DOUTOR EM INFORMÁTICA
ESPECIALIDADE ENGENHARIA INFORMÁTICA

pela

FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DE LISBOA
DEPARTAMENTO DE INFORMÁTICA

Tese orientada pelo Prof. Doutor Luís Manuel Pinto da Rocha Afonso Carriço

2011

To my dearest beloveth.

To my parents, Gabriela and Eduardo, for making me what I am today.

Acknowledgements

The biggest acknowledgement goes, without a doubt, to my fellow advisor, professor Luís Carriço. He has proven to have the deepest insights and thoughtful comments and advices towards the success of my research scientist career. A big, big **thank you**!

I would like to thank also to my colleagues and fellow researchers from the AbsInt group, from the HCIM research line, as well as from the LaSIGE research laboratory, for keeping up with all my discussions, bursts, chatting, "sweat & tears", and general blabber during the last years. This includes the Portuguese Web Archive team at FCCN, for which without their efforts, availability, discussions, data, and computing platform, several parts of my experimental work difficultly would be accomplished.

A special thanks goes also to all the anonymous reviewers that patiently replied with thoughtful – oft thought-provoking – insights and comments in all manuscripts I submitted to conferences and journals. It definitely made my work stronger, more detailed, providing me with food for thought to transform negative reviews into positive forces to improve my research. Another special thank to Carlos Duarte for his thorough review of this thesis' texts. A particular heads up goes to the guys at the ACM W4A (Web4All) conference, which proved to be a special place and research community for the work conducted during my research career.

Finally, but not the least, all of the work described in this thesis would be definitely hard to achieve without the financial support provided by Fundação para a Ciência e Tecnologia (scholarship SFRH/BD/29150/2006), for which I am grateful.

Abstract

Accessibility quality on the Web is essential for providing a good Web experience to people with disabilities. The existence of *virtual ramps* aid these users grasping and interacting with Web content, just like the experience of those who are unimpaired. However, more often than not, Web pages impose accessibility barriers, usually centred on the unavailability of tailored content to specific perceptual abilities (e.g., textual description of images, enabling grasping information with assistive technologies), as well as on proper HTML structural elements that adequate the semantics of a Web page.

When evaluating the accessibility quality of Web pages, the resulting analysis is often focused on a small sample set (e.g., a single Web page or a selection of pages from a Web site). While this kind of analysis gets the gist of accessibility quality, it misses the big picture on the overall accessibility quality of the Web.

This thesis addresses the challenge of observing accessibility phenomena on the Web, through the experimental evaluation of large collections of Web pages. This resulted on new findings about the accessibility quality of the Web, such as its correlation with HTML element count, and the erroneous perception of accessibility quality by developers. Small-scale experiments have been verified also at large scale, such as the correlation between the usage of HTML templates and accessibility quality. Based on the challenges raised by the experimental evaluation, this thesis proposes a novel approach for large scale Web accessibility evaluation based on Linked Data, as well as the establishment of metrics to assess the truthfulness and coverage of automated evaluation methods.

Keywords: Web Accessibility, Web Science, Web Characterisation

Resumo - Portuguese Abstract

A qualidade da acessibilidade é um factor crucial para as pessoas com deficiências terem uma boa experiência de interacção com a Web. A existência de *rampas virtuais* ajuda estas pessoas a compreender e interagir com conteúdos Web, a par do que o utilizador comum já experiencia. Porém, a maioria das páginas Web ainda contêm barreiras à acessibilidade. Estas barreiras centram-se normalmente na indisponibilidade de conteúdos perceptíveis por diferentes tipos de capacidades (e.g., descrições textuais de imagens), bem como no uso incorrecto de elementos HTML de acordo com a semântica de uma página Web.

Nos dias de hoje, a avaliação da qualidade de acessibilidade de páginas Web é ainda efectuada em pequena escala (e.g., uma página Web ou, no melhor caso, um conjunto de páginas representativas de um sítio Web). Apesar deste tipo de avaliações resultarem na compreensão de alguns fenómenos do estado da acessibilidade na Web, ainda não se sabe qual o seu impacto em larga escala.

Esta tese discute os principais desafios na observação da acessibilidade da Web, tendo por base um conjunto de avaliações experimentais de colecções de grande dimensão de páginas Web. Destes estudos destacam-se as seguintes contribuições e resultados:

- **A diferença drástica na interpretação dos *avisos* resultantes de avaliações de acessibilidade Web:** um dos resultados principais da avaliação experimental em larga escala destaca a diferença na interpretação dos *avisos* (*warnings*) da aplicação de técnicas da norma WCAG, onde a interpretação optimista (i.e., a visão da maioria dos criadores de páginas Web) se distancia amplamente da interpretação conservadora (onde os *avisos* são interpretados como *erros*);

- **A correlação entre a qualidade da acessibilidade de uma página Web e a sua complexidade:** este mesmo estudo de larga escala revelou uma correlação entre a complexidade de uma página Web (no que diz respeito ao número de elementos HTML que contém) e a qualidade da acessibilidade. Quanto menor a complexidade de uma página Web, mais certa se torna a alta qualidade da acessibilidade dessa página;
- **O benefício do uso de *templates* e sistemas de gestão de conteúdos na melhoria da acessibilidade de páginas Web**: em ambos os estudos experimentais de acessibilidade foi detectada uma correlação entre a qualidade de acessibilidade das páginas Web e o uso de *templates* e sistemas de gestão de conteúdo. Esta propriedade foi verificada quer em pequena escala (sobre uma colecção de páginas Web da *Wikipedia*), quer em larga escala;
- **O incumprimento das regras mais elementares e mais conhecidas da acessibilidade:** estes estudos experimentais permitiram também verificar que, apesar de toda a envagelização e educação sobre as questões de acessibilidade na Web, a maioria das regras de acessibilidade são incessantemente quebradas pela maioria das páginas Web. Esta problemática verifica-se, em particular, nas regras de cumprimento de acessibilidade mais conhecidas, tal como por exemplo a disponibilidade de textos alternativos a conteúdos multimédia.

Com base nestas experiências e resultados, esta tese apresenta um novo modelo de estudo da acessibilidade na Web, tendo por base o ciclo de estudos da Web em larga escala. Deste modelo resultaram as seguintes contribuições:

- **Um modelo para a avaliação distribuída de acessibilidade Web, baseado em propriedades *tecnológicas* e *topológicas*:** foi concebido um modelo de avaliação de acessibilidade Web que permite a concepção de sistemas de avaliação com base em propriedades tecnológicas e topológicas. Este modelo possibilita, entre outras características, o estudo da cobertura de plataformas e avaliadores de acessibilidade, bem como da sua aplicação em larga escala;

- **Uma extensão às linguagens e modelos EARL e Linked Data, bem como um conjunto de definições para extrair informação destes:** este modelo de avaliação de acessibilidade Web foi sustentado também pela sua concretização em linguagens e modelos já existentes para o estudo de acessibilidade (EARL) e da Web em larga escala (Linked Data), permitindo assim a sua validação;
- **Definição dos limites da avaliação de acessibilidade Web:** por fim, este modelo de avaliação de acessibilidade permitiu também delinear uma metodologia de meta-avaliação da acessibilidade, na qual se poderão enquadrar as propriedades dos avaliadores de acessibilidade existentes.

Todas estas contribuições resultaram também num conjunto de publicações científicas, das quais se destacam:

- **Rui Lopes** and Luís Carriço, *A Web Science Perspective of Web Accessibility*, in submission for the ACM Transactions on Accessible Computing (TACCESS), ACM, 2011;
- **Rui Lopes** and Luís Carriço, *Macroscopic Characterisations of Web Accessibility*, New Review of Hypermedia and Multimedia – Special Issue on Web Accessibility. Taylor & Francis, 2010;
- **Rui Lopes**, Karel Van Isacker and Luís Carriço, *Redefining Assumptions: Accessibility and Its Stakeholders*, The 12th International Conference on Computers Helping People with Special Needs (ICCHP), Vienna, Austria, 14-16 July 2010;

- **Rui Lopes**, Daniel Gomes and Luís Carriço, *Web Not For All: A Large Scale Study of Web Accessibility*, W4A: 7th ACM International Cross-Disciplinary Conference on Web Accessibility, Raleigh, North Carolina, USA, 26-27 April 2010;
- **Rui Lopes**, Konstantinos Votis, Luís Carriço, Dimitrios Tzovaras, and Spiridon Likothanassis, *The Semantics of Personalised Web Accessibility Assessment*, 25th Annual ACM Symposium on Applied Computing (SAC), Sierre, Switzerland, 22-26 March, 2010
- Konstantinos Votis, **Rui Lopes**, Dimitrios Tzovaras, Luís Carriço and Spiridon Likothanassis, *A Semantic Accessibility Assessment Environment for Design and Development for the Web*, HCI International 2009 (HCII 2009), San Diego, California, USA, 19-24 July 2009
- **Rui Lopes** and Luís Carriço, *On the Gap Between Automated and In-Vivo Evaluations of Web Accessibility*, HCI International 2009 (HCII 2009), San Diego, California, USA, 19-24 July 2009;
- **Rui Lopes**, Konstantinos Votis, Luís Carriço, Spiridon Likothanassis and Dimitrios Tzovaras, *Towards the Universal Semantic Assessment of Accessibility*, 24th Annual ACM Symposium on Applied Computing (SAC), Waikiki Beach, Honolulu, Hawaii, USA, 8-12 March 2009;
- **Rui Lopes** and Luís Carriço, *Querying Web Accessibility Knowledge from Web Graphs*, Handbook of Research on Social Dimensions of Semantic Technologies, IGI Global, 2009;
- **Rui Lopes**, Konstantinos Votis, Luís Carriço, Spiridon Likothanassis and Dimitrios Tzovaras, *A Service Oriented Ontological Framework for the Semantic Validation of Web Accessibility*, Handbook of Research on Social Dimensions of Semantic Technologies, IGI Global, 2009;

- **Rui Lopes** and Luís Carriço, *On the Credibility of Wikipedia: an Accessibility Perspective*, Second Workshop on Information Credibility on the Web (WICOW 2008), Napa Valley, California, USA, 2008;
- **Rui Lopes**, Luís Carriço, *A Model for Universal Usability on the Web*, WSW 2008: Web Science Workshop, Beijing, China, 22 April 2008;
- **Rui Lopes**, Luís Carriço, *The Impact of Accessibility Assessment in Macro Scale Universal Usability Studies of the Web*, W4A: 5th ACM International Cross-Disciplinary Conference on Web Accessibility, Beijing, China, 21-22 April 2008. **Best paper award**;
- **Rui Lopes**, Luís Carriço, *Modelling Web Accessibility for Rich Document Production*, Journal on Access Services 6(1-2), Routledge, Taylor & Francis Group, 2009;
- **Rui Lopes**, Luís Carriço, *Leveraging Rich Accessible Documents on the Web*, W4A: 4th ACM International Cross-Disciplinary Conference on Web Accessibility, Banff, Canada, 7-8 May 2007.

Palavras Chave: Accessibilidade Web, Ciência da Web, Caracterização da Web

Contents

List of Figures

List of Tables

Code Listings

There's real poetry in the real world. Science is the poetry of reality.

Richard Dawkins

Chapter 1

Introduction

1.1 Overview

The idea of interconnecting information aided by technology was first idealised by Bush (1945), with the proposal of a device named *Memex*. This theoretical device was based on the notion of establishing relationships between different sources of information stored in microfilms. Later on, computer based systems inspired by Memex began to appear, such as the oNLine System by Engelbart & English (1968), the Hypertext Editing System by Carmody *et al.* (1969), and Xanadu by Nelson (1980). However, these academic systems posed too much effort on content creation and maintenance, and were based on primitive technology. All of these factors provided the opportunity for the World Wide Web to flourish. Its pragmatic easiness of use and openness leveraged it as the de-facto hypertext/hypermedia system, which was quickly embraced throughout the world.

The invention of the World Wide Web by Berners-Lee *et al.* (1992) has changed the way people consume and produce information for different purposes. Since its inception, it has been used as a way to publish interconnected information through *hyperlinks*. Special purpose *User Agents* such as Web browsers gave users an easy way to navigate through information presented on a computer screen with point-and-click interactions, as well as simple authoring methods to create contents and make them available online.

The simplicity of publishing and consuming information on the Web has resulted, throughout the years, in a dramatic increase (both in size and complexity) of available content. Websites have been created to fulfill different needs and activities for different kinds of users: information gathering, communication, transactions, entertainment, etc. With the creation of more powerful supportive technologies and richer Web browsers, the characteristics of the Web have made it possible to expand beyond its traditional one-way information channel capabilities. This has resulted on viewing the Web as a living organism with millions of nodes (i.e., Web pages) and, ergo, an immense diversity.

As de Kunder (2010) shows, current estimates of the size of the Web result on the availability of 45 billion Web pages, growing at a fast pace everyday (without taking into account the deep Web, i.e., Webpages that are not reachable through automatic methods). In comparison, the human brain has roughly 100 billion neurons, which puts the Web at the same order of magnitude of complexity. Therefore, it is a challenge to understand and discover what order lays from the Web's content and structure at different abstraction levels.

This study of the Web as a first-order entity is the goal of the recently created Web Science discipline (Berners-Lee *et al.*, 2006). It focuses on how Web pages are constructed, how links are formed between them, and how this structure is leveraged in different directions. One of the outcomes from this approach is studying how people interact with the Web. This interaction capability not only concerns end users, i.e., those who browse and search content, purchase goods online, and use this medium for entertainment and work, but also those users who create content and all of the underlaying piping that sustains the Web (e.g., developers, designers).

The Web presents different challenges to users, specially regarding the action they have to perform to fulfil their goals. Factors such as user characteristics and interaction mediation devices such as the mouse or the keyboard, are fundamental pillars of the *context* upon which a user interacts with the Web, which might pose severe challenges on interaction activities with the Web. With the ever-increasing number of people accessing the Web anytime, anywhere, a high number of less common (and unexpected) situations begin to have relevance. Examples include a blind person's inability to perceive and interact with a given content, or the difficulty of navigating through a website on a mobile phone. Such situations pose severe consequences on the satisfaction levels of users while trying to interact with websites, leading them to find more adequate and pleasuring alternatives elsewhere on the Web (Li *et al.*, 2006).

This satisfaction level is a well known problem studied by the Human-Computer Interaction and Experience Design research fields. *User Experience*, as it is generally denominated, centres on *"all aspects of the user's interaction with the product: how it is perceived, learned, and used"*, as defined by Norman (1998). While aesthetics do have a role on increasing satisfaction levels, *usability* (i.e., ease of use) and *accessibility* (i.e., ability to access) have a profound influence on user experience (Norman, 1998). These two concepts establish the levels upon which the user's goals on interacting with a website are completely fulfilled as expected.

It is important to understand that *user experience* is regarded differently by each audience, based on their (dis)abilities, constraints, and expectations.

When taking into account both usability and accessibility factors in an integrated way, the notion of *Universal Usability* materialises. As described in Shneiderman (2000), it is one of the key challenges for Web Science (Shneiderman, 2007). This challenge is centred on exploring the accommodation of software (and their user interfaces, in particular) to both technology variety, user diversity and user knowledge, instead of providing a one-size-fits-all solution. Usability and accessibility are, therefore, the root for providing a high quality user experience in software, including the Web.

As explained, the user's entry point of interacting with the Web is the browser. It is the browser's responsibility to render the data transmitted by a Web server through the HTTP protocol (Fielding *et al.*, 1998), and managing all interaction events triggered by users. This data, which ultimately composes a Web page – *the* front-end –, is defined through a mix of technologies interpreted by the browser, including HTML (Pemberton *et al.*, 2002), CSS (Bos *et al.*, 2007), and JavaScript and associated event models (Höhrmann *et al.*, 2006).

These three technologies have a deep impact on all tasks performed over a Web front-end, since they dictate how information is presented to the user, as well as how it should be interacted. Therefore, the analysis of the quality of the Web from a universal usability perspective must take into account that these Web front-end technologies have to be analysed from the perspective of the specific characteristics of each target audience: users, devices, and environment. This fact has the consequence that, for users with disabilities, front-ends should comprise accessibility features so that content, navigation, and interaction are not hindered.

1.2 Problem

It has been recurrently observed that the expansion of the Web is going beyond the highest percentile of users and situations: unimpaired users interacting with the Web on a desktop computer (c.f. Harper & Yesilada (2008)). The spectrum of possibilities is opening the way to new interaction capabilities to consume and produce contents to the Web. The increasing diversity of users and devices, as well as of publishing, information sharing, and communication platforms is shaping the Web in multiple ways.

The introduction of end users as content producers brings another variable in what respects to the adequacy of Web pages to all users. Coping with such a rich, complex scenario is a big challenge for multidisciplinary Web development teams, including designers, usability and accessibility experts, and managers, besides developers.

However, as of now, teams typically lean towards providing a high user experience to the aforementioned percentile, nothing else. Most of the times the target audiences for products and services put on the Web correspond to this. But, through the correct application of Web front-end technologies, a primary degree of user experience quality to users can be delivered. For instance, by carefully crafting CSS directives, Web front-ends can cope with different screen configurations.

But creating front-ends that afford the capability limitations of users with disabilities requires additional efforts from Web development teams. Thus, in an ideal scenario, when thinking of accessibility issues, teams can follow WCAG, the Web Content Accessibility Guidelines (Caldwell *et al.*, 2008), a widely accepted standardisation effort on how to create accessible Web pages, as specified by the World Wide Web Consortium[1]. By following these guidelines, teams can have a certain degree of confidence that a single front-end can be perceived, interacted, and experienced both by unimpaired users and those with disabilities in a similar way. But, more often than not, teams do not have the expertise, motive, the right tools, or even time to cope with making Web sites accessible (Lopes *et al.*, 2010b).

All of the impedance between user needs and compliance from developers and designers raises a set of problems, synthesised as follows:

> **Problem:** Developers and designers often lack the awareness and skills to create accessible Web pages. As content production and consumption becomes richer, accessibility issues are increasingly problematic. Currently, little is known about the corresponding impact

[1] http://www.w3.org/

on a large scale and how it relates with the microscopic perspectives of developers and designers.

This broad perspective emerging from large scale studies is fundamental to boost the aimed awareness and find focused and productive approaches for accessible Web development. Further subdividing this problem, more detailed issues arise, as follows.

1.2.1 Users and the Inaccessible Web

P1: It has been thoroughly studied, e.g., in Harper & Yesilada (2008), that a lot of content on the Web has accessibility barriers. However, how this lack of accessibility is spread on the Web has yet to be known.

1.2.2 Developers' Unawareness

P2:Accessibility problems on the Web are often related to developers and designers' unawareness or distorted understanding of accessibility and inclusive practices.

1.2.3 End-user Content Production

P3: End-user content production on the Web (e.g., wikis, blogs) is often supported by content management systems (CMSes) that hide the complexity of Web technologies. However, that support is still no guarantee to deliver truthfully accessible content.

1.2.4 Large-scale Accessibility Evaluation

P4: Accessibility evaluation is mostly applied in small scales (e.g., a single Web site) through manual or semi-automated procedures, since tools are targeted at developers and designers. This leads to a lack of adequacy from current evaluation methods to the complexity and scale of evaluating large portions of the Web.

1.3 Hypothesis

Based on the problems elicited in the previous Section, this thesis' focus pertains on the measurement and analysis of Web accessibility at a large scale, in order to understand the influence of the Web on content accessibility for people with disabilities and, at the same time, how developers and designers weave accessibility onto the Web. Thus, this work proposes the following research hypothesis:

> *Web accessibility evaluation at large scale will leverage new knowledge about the Web itself, its development and adequate mechanisms to raise proper awareness of designers and developers towards accessibility.*

The work presented in this thesis has been devised in order to validate this hypothesis. This has been done by devising a set of experimental studies, which have been reflected onto a theoretical model for Web accessibility evaluation at large scale.

1.4 Contributions

The work presented in this thesis resulted in contributions on *theoretical* models for Web accessibility evaluation, *experimental* analysis of the state of accessibility quality on the Web and accompanying ancillary *software and data*, all of which have been supported by peer-reviewed scientific *publications*. Another contribution concerns the connecting points between the thesis and ongoing *research projects*. All of these are presented next.

1.4.1 Theoretical

The ground foundation of this thesis lays at how to understand Web accessibility at large scale. On this basis, this work contributes to the following theoretical advances:

- A model for distributed Web accessibility evaluation, based on *technological* and *topological* properties;

- A reformulation of accessibility evaluation for large scale analysis, and how accessibility knowledge can be mined through Linked Data principles.

1.4.2 Experimental

On the experimental side, the work described in this thesis provides a new set of insights on how accessibility is shaped on the Web, including:

- The drastic difference in interpretation of *warnings* in Web accessibility evaluation results by developers and designers. These stakeholders' perception of an evaluation *warning* being less important than an evaluation *error* leads to the dismissal of the former (i.e., less important or critical), contributing to a worsened quality of accessibility on the Web, even when supported by accessibility evaluation tools;

- The correlation between the accessibility quality variation of a Web page and its complexity (in terms of number of HTML elements). The conducted large scale experiment leveraged knowledge that the current state of accessibility on the Web is that smaller Web pages tend to have better accessibility quality;

- The usage of templates and content management systems (CMSes) to improve the accessibility quality of a Web page. A study developed over Wikipedia analysed a set of internal Web pages and external Web pages linked from them (used for information credibility purposes). This Web site leverages different template and CMSes mechanisms that, according to this experimental study, resulted in a stable, high-quality accessibility compliance. This contrasts with the externally linked pages, where accessibility variability is higher. As a corollary of this study, these templates and CMSes should provide additional support for accessibility, since content producers still have to comply with accessibility features not automatically covered by these systems;

- The lack of compliance with the simpler and most well known accessibility guidelines (i.e., alternative textual content to describe pictorial content).

> Despite all the evangelism and education towards the perception of accessibility issues, accessibility is still left as an afterthought in most cases on the Web.

1.4.3 Software and Data

During the course of this work, some software components for Web accessibility evaluation and analysis were developed for the conducted experiments. These also resulted in the generation of different data sets comprising of Web accessibility evaluation data points. The following compilation lists these components and data sets:

- *UUCrawler*. A set of software components to analyse the accessibility of Web pages, particularly centred on the URL structure and page content of Wikipedia.
  - *Evaluator*. This component is responsible for the analysis of a Web page HTML structure and content, from an accessibility point-of-view. It implements an automatable subset of WCAG 1.0;
  - *Crawler*. This component crawls a random selection of Web pages from Wikipedia, as well as external Web pages marked as citations or references on each Wikipedia article;
  - *Data set*. The resulting data set from the experimental analysis of Wikipedia, comprising a total 365 Web pages (internal and external) evaluated.
- *QualWeb*. A set of software components for the analysis of accessibility quality of large collections of Web pages, with a strong emphasis on their application on distributed architectures for information processing.
  - *Evaluator*. This main component is responsible for a fast analysis of the accessibility quality inherent of the HTML structure and content of Web pages. It implements an automatable subset of WCAG 1.0, extending *UUCrawler* with new, improved analysis;

- *Hadoop Component.* A wrapper of the *evaluator* component that can be plugged into a Hadoop cluster, configured according to the distributed architecture devised by the Portuguese Web Archive team;
- *Linked Data Component.* A wrapper of the *evaluator* component that provides evaluation results of Web accessibility conforming to *Linked Data* principles;
- *Interactive UI Component.* An application built on top of the *evaluator* component that provides an interactive service for accessibility evaluation of Web pages, as well as reporting, analysis, and repair features;
- *Data set.* The resulting data set from the experimental analysis of a large page collection from the Portuguese Web Archive Initiative, comprising the evaluation of almost 30 million Web pages.

Work is being conducted on making available the software components through an open-source license, as well as making available the data sets for further research initiatives.

1.4.4 Publications

The work presented in this thesis has been peer-reviewed through scientific publications, including conference papers, journal articles, and book chapters. The following compilation of publications is directly or indirectly related to the goals of this thesis:

- **Rui Lopes** and Luís Carriço, *A Web Science Perspective of Web Accessibility*, in submission for the ACM Transactions on Accessible Computing (TACCESS), ACM, 2011. This work presents the model for distributed Web accessibility evaluation at large scale;

- **Rui Lopes** and Luís Carriço, *Macroscopic Characterisations of Web Accessibility*, New Review of Hypermedia and Multimedia – Special Issue on Web Accessibility. Taylor & Francis, 2010. This work expands on the results of the large-scale Web accessibility evaluation performed over the Portuguese Web, with an emphasis on hyperlinking properties;

- **Rui Lopes**, Karel Van Isacker and Luís Carriço, *Redefining Assumptions: Accessibility and Its Stakeholders*, The 12th International Conference on Computers Helping People with Special Needs (ICCHP), Vienna, Austria, 14-16 July 2010. This work describes a survey of several accessibility stakeholders, in order to understand the state-of-the-art of accessibility compliance and corresponding effects on people with disabilities;

- **Rui Lopes**, Daniel Gomes and Luís Carriço, *Web Not For All: A Large Scale Study of Web Accessibility*, W4A: 7th ACM International Cross-Disciplinary Conference on Web Accessibility, Raleigh, North Carolina, USA, 26-27 April 2010. This work presents the first results of a large-scale Web accessibility evaluaton performed over the Portuguese Web;

- **Rui Lopes**, Konstantinos Votis, Luís Carriço, Dimitrios Tzovaras, and Spiridon Likothanassis, *The Semantics of Personalised Web Accessibility Assessment*, 25th Annual ACM Symposium on Applied Computing (SAC), Sierre, Switzerland, 22-26 March, 2010. This work details how Web accessibility evaluation reports can be mined in personalisation scenarios;

- Konstantinos Votis, **Rui Lopes**, Dimitrios Tzovaras, Luís Carriço and Spiridon Likothanassis, *A Semantic Accessibility Assessment Environment for Design and Development for the Web*, HCI International 2009 (HCII 2009), San Diego, California, USA, 19-24 July 2009. This work presents a semantics-centric framework for Web accessibility development and evaluation;

- **Rui Lopes** and Luís Carriço, *On the Gap Between Automated and In-Vivo Evaluations of Web Accessibility*, HCI International 2009 (HCII 2009), San Diego, California, USA, 19-24 July 2009. This study details the limits between automated evaluations of accessibility and how these can be correlated with user studies;

- **Rui Lopes**, Konstantinos Votis, Luís Carriço, Spiridon Likothanassis and Dimitrios Tzovaras, *Towards the Universal Semantic Assessment of Accessibility*, 24th Annual ACM Symposium on Applied Computing (SAC), Waikiki Beach, Honolulu, Hawaii, USA, 8-12 March 2009. This work presents a set of requirements elicited from user studies that must be taken into account for Web accessibility evaluation;

- **Rui Lopes** and Luís Carriço, *Querying Web Accessibility Knowledge from Web Graphs*, Handbook of Research on Social Dimensions of Semantic Technologies, IGI Global, 2009. This work details how accessibility knowledge can be mined through Linked Data principles;

- **Rui Lopes**, Konstantinos Votis, Luís Carriço, Spiridon Likothanassis and Dimitrios Tzovaras, *A Service Oriented Ontological Framework for the Semantic Validation of Web Accessibility*, Handbook of Research on Social Dimensions of Semantic Technologies, IGI Global, 2009. This work presents how Web accessibility evaluation can be provided through Web services, in order to facilitate *distributed* evaluation scenarios;

- **Rui Lopes** and Luís Carriço, *On the Credibility of Wikipedia: an Accessibility Perspective*, Second Workshop on Information Credibility on the Web (WICOW 2008), Napa Valley, California, USA, 2008. This paper presents a study on the limits of template-centric CMSes such as Wikipedia, and how these can pose problems on information credibility;

- **Rui Lopes**, Luís Carriço, *A Model for Universal Usability on the Web*, WSW 2008: Web Science Workshop, Beijing, China, 22 April 2008. This work details the first insights for a Web Science perspective on Web accessibility evaluation and reporting;

- **Rui Lopes**, Luís Carriço, *The Impact of Accessibility Assessment in Macro Scale Universal Usability Studies of the Web*, W4A: 5th ACM International Cross-Disciplinary Conference on Web Accessibility, Beijing, China, 21-22 April 2008. This work details a first experimental study of Web accessibility evaluation, based on Wikipedia Web pages. **Best paper award**;

- **Rui Lopes**, Luís Carriço, *Modelling Web Accessibility for Rich Document Production*, Journal on Access Services 6(1-2), Routledge, Taylor & Francis Group, 2009. This paper further details a set of accessibility modelling capabilities for Web pages, in order to understand the limits of accessible Web design and corresponding compliance;

- **Rui Lopes**, Luís Carriço, *Leveraging Rich Accessible Documents on the Web*, W4A: 4th ACM International Cross-Disciplinary Conference on Web Accessibility, Banff, Canada, 7-8 May 2007. This paper presents a first set of accessibility modelling capabilities for Web pages, which takes into account template-based mechanisms as a way to improve accessibility features.

1.4.5 Research Projects

The bulk of the work presented in this thesis has been conducted through a PhD scholarship funded by Portugal's National Science Foundation (FCT). Notwithstanding, during the course of this work, it has also been partially associated with some research projects that tackle different aspects of accessibility:

- *ACCESSIBLE: Accessibility Assessment Simulation Environment for New Applications Design and Development.* This European Commission FP7 ICT project centres on the design and development of accessibility evaluation and simulation tools. It aims at helping accessibility stakeholders – particularly designers and developers – creating applications that are more accessible for people with disabilities, in different application domains: *Web*, *Mobile*, *Web Services*, and *Description Languages*. The consortium of this project includes Oracle Corporation (formerly Sun Microsystems), CERTH, FFCUL, Marie Curie Association, amongst other partners.

  The key points between ACCESSIBLE and the work presented in this thesis concern the analysis of accessibility standards and stakeholders. However, the focus of the project lays at improving the design and development process for accessible applications, where this thesis discusses new findings about Web accessibility. These findings will, ultimately, be fed into the ACCESSIBLE project as additional information to improve the developed evaluation and simulation tools.

  More information about the ACCESSIBLE project can be found at http://www.accessible-project.eu/.

- *QualWeb*. This FCT national research project is focused on the analysis of large Web document collections, from an accessibility point-of-view. QualWeb aims at characterising different accessibility issues of the Portuguese Web, such as its adequacy to specific disabilities. This project is a joint collaboration between FFCUL and FCCN, through the Portuguese Web Archive Initiative.

  This project entails on the initial findings of Web accessibility characterisation at large scale that resulted from the work presented in this thesis. It is expected that the ongoing research on Web accessibility evaluation discussed in this thesis will continue during the course of the QualWeb project, especially on tackling the future challenges presented later on in this document.

  More information about the QualWeb project can be found at http://hcim.di.fc.ul.pt/wiki/QualWeb.

1.5 Thesis Structure

This rest of this thesis is organised as follows: *Chapter 2* presents a set of background concepts on the Web science and Web accessibility domains; *Chapter 3* details the state-of-the art of Web accessibility guidelines and evaluation, as well as an overview of Web Science particularly focused on Web topology analysis; afterwards, *Chapter 4* details experimental studies on the analysis of the accessibility of Web page collections, with the goal of finding new characteristics of accessibility on the Web; based on results from these studies, *Chapter 5* presents a model for large-scale evaluations of accessibility. In *Chapter 6*, the QualWeb evaluator is presented, with the focus on the role it plays in both the experimental studies and distributed accessibility evaluations. Finally, *Chapter 7* revisits the hypothesis raised earlier in this Chapter, in the light of the results presented in this thesis, and briefly discusses the future directions of Web accessibility evaluation.

The power of the Web is in its universality. Access by everyone regardless of disability is an essential aspect.

Tim Berners-Lee

Chapter 2

Background Concepts

2.1 Introduction

As presented in the previous Chapter, the focus of this thesis lays at the study of the World Wide Web and its adequacy to people with disabilities. The work described hereinafter makes several assumptions based on how the Web is architected, how people with disabilities interact with computers, as well as how other accessibility stakeholders (e.g., developers) coexist within this scenario. Hence, this Chapter details the background knowledge and corresponding concepts about the Web, its architecture and functioning, as well as on the accessibility of information through computers and what goals and challenges are posed to accessibility stakeholders.

2.2 The World Wide Web

2.2.1 Architecture

People use the Web by navigating from link to link, in order to meet their goals (Kellar *et al.*, 2006). While each goal and each person reflects on different navigating strategies (e.g., starting with a query on a search engine, following a bookmark, etc.), the basic interaction premise lays on the consecutive HTTP (Fielding *et al.*, 1998) *Request-Response* messages exchange between a *User agent* and a *Web server* (Figure 2.1). More often than not, each response is in the form of a Web page, conveying information and hyperlinks to other Web pages.

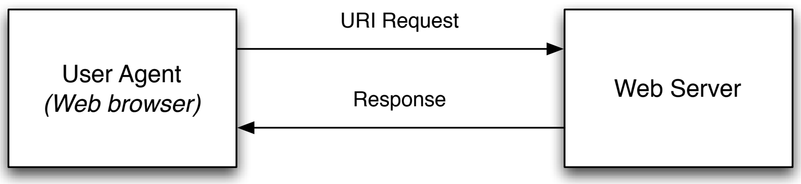


Figure 2.1: *Request-Response* message exchange on the Web

The HTTP protocol defines *User agents* as client applications that interact with Web servers. This communication is made with HTTP request methods such as *GET*, *POST*, etc. Each request received by a Web server is processed, and an HTTP response is sent back to the user agent, depending on the semantics specified by the application running on the Web server. This simple client-server communication lays at the core of all HTTP user agents, the most important one being the *Web browser*. Other user agents include *Web crawlers*, used by search engines for a programmatic browsing of the structure of the World Wide Web.

This interaction with Web resources (such as Web pages) between Web servers and user agents lays on top of the basic elements of the architecture of the Web (Jacobs & Walsh, 2004), as depicted in Figure 2.2: *URIs*, *Resources*, and *Representations*.

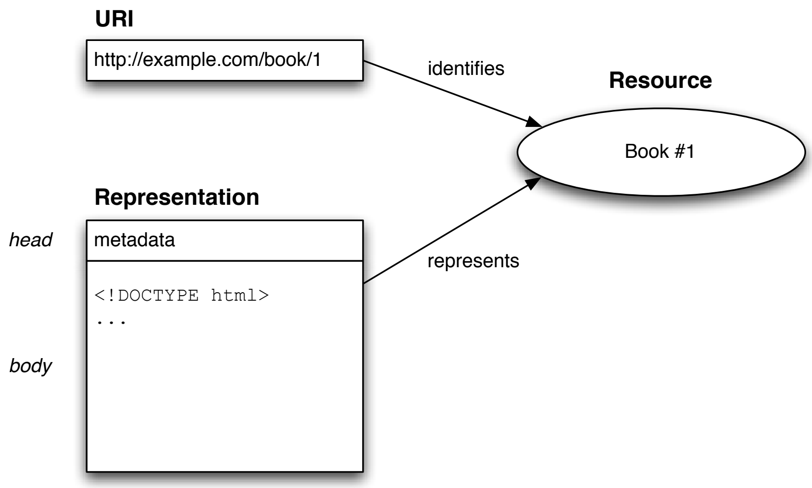


Figure 2.2: Elements of the Architecture of the Web

Each request made by a User agent is in the form of a *URI* (Berners-Lee *et al.*, 2005). Each URI identifies a *resource* that may be made available by a Web server (if the URI is, in fact, a URL), ideally in a *representation* that can be parsed/interpreted by the User agent – such as HTML (Pemberton *et al.*, 2002) in the case of Web browsers. The simplicity of these basic elements has spurred the production of a wealth of information that defines the Web – a thorough, interactive, navigable graph of content represented by Web pages.

2.2.2 Semantic Web

The Semantic Web Activity at the W3C[1] was created with the aim of providing a common knowledge modelling framework, particularly focused on data sharing and reuse across applications and communities. Just like the World Wide Web was defined as a Web of documents, the Semantic Web aims at being a *Web of Data* that can be mined for information and knowledge in different ways. By leveraging the uniqueness properties of URIs, the Semantic Web is decomposed into a set of abstraction and knowledge modelling layers, as depicted in Figure 2.3:

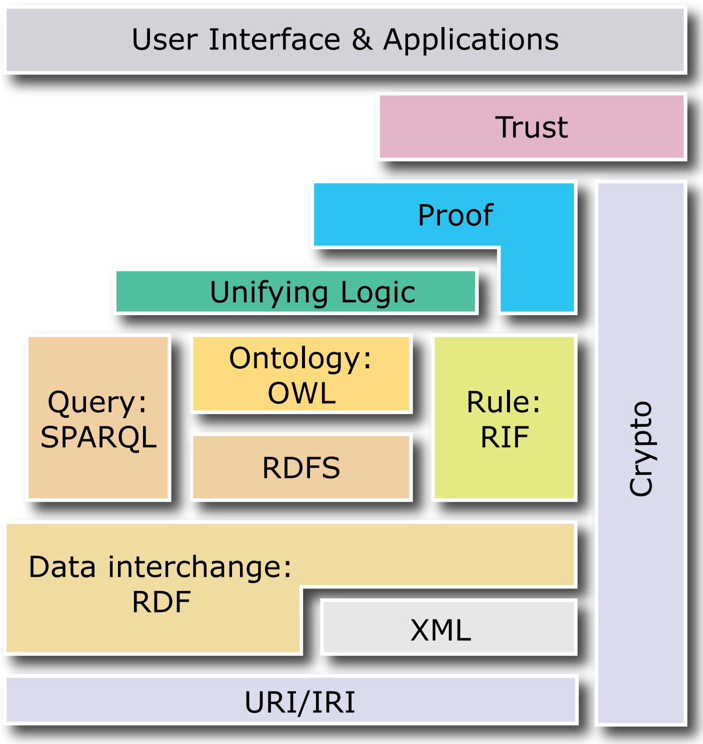


Figure 2.3: Elements of the Semantic Web

The Resource Description Framework (RDF) (Beckett & McBride, 2004) has been created to overcome the information ambiguity problem, by allowing the univocal, semantic description and modelling of information. For this purpose, RDF leverages the architecture of the Web through the use of URIs to identify

[1]Semantic Web Activity: http://www.w3.org/2001/sw/

information in the form of *Resource-Property-Literal* triples (Manola & Miller, 2004). An example RDF is presented below, containing two triples:

- The *resource* represented at http://en.wikipedia.org/wiki/Tim_Berners-Lee has a title *property* (i.e., `dc:title`) with a *literal* value of *Tim Berners-Lee*;
- The same *resource* was *published* in *Wikipedia*.

```
@prefix dc: <http://purl.org/dc/elements/1.1/>.

<http://en.wikipedia.org/wiki/Tim_Berners-Lee>
  dc:title "Tim Berners-Lee";
  dc:publisher "Wikipedia".
```

Listing 2.1: An example of RDF (in the Notation3 serialisation format)

Such triples afford a rich expressivity of knowledge that effectively form a *Web of data*. On top of RDF, standards such as RDFS (Brickley & Guha, 2004) and OWL (Schreiber & Dean, 2004) afford a more thorough description of resource semantics (e.g., classes of concepts). All information coded through RDF can then be processed through machine-only methods such as SPARQL (Prud'hommeaux & Seaborne, 2008) or even through human interaction with special browsers (Berners-lee *et al.*, 2006). Other elements and layers of the Semantic Web include RIF (Boley *et al.*, 2010) – Rules Interchange Format – and *Unifying Logic* and *Proof* (these latter two have not been specified yet) to build a formal domain of analysis for logic reasoning and inference. Finally, *Trust* (one of the main goals to be achieved with the Semantic Web), builds on top of both *Logic* and *Crypto* concepts, so that truthful data can be formed in this *Web of data*. While all of these layers are of the uttermost importance within the *Semantic Web*, the work presented in this thesis focuses on implementing the lower layers (URI, RDF, OWL, and SPARQL) in the accessibility domain.

2.2.3 Linked Data

Despite being based on URIs, the Semantic Web framework does not promote the interchange and interconnection of disparate data sources. Consequently, most data created according to the Semantic Web principles has been done within information silos (e.g., closed, local systems, outside the reach of the Web). One of the reasons for this concerns the fact that URIs might not be dereferenced into real URLs (i.e., they are just identifiers, not recognised nor reachable on the Web itself). This issue led to a disparate, disconnected amount of information codified in RDF and related Semantic Web technologies. To mitigate these limitations, the *Linked Data Initiative*[1] appeared as an answer to foster the organic growth of the *original* Web (i.e., Web pages and hyperlinks) into Semantic Web concepts.

Consequently, *Linked Data* has been proposed by Berners-Lee (2009) as a small set of steps to tailor RDF information into a navigable form (both by machines and humans). With the surge of linked data, several interconnected RDF datasets have emerged, as exemplified in Figure 2.4. Canonical examples include DBpedia[2], a machine-processable counterpart of the knowledge concepts and relationships contained within Wikipedia[3].

2.3 Accessibility

User interfaces – the *visible* side of computer applications – are created, as the name implies, as a way for humans to interact with the processes defined in software. Each user decides a set of tasks to be performed over user interfaces, in order to achieve a specific goal, such as filling a *form* in order to purchase a good in an online shop, or browsing information for research purposes. The way these user interfaces are adequate to users' needs and goals is the main research object of the Human-Computer Interaction (HCI) research field. This includes studying how user-centred application design helps creating user interfaces that increase productivity and experience for users.

One of the key concepts from the HCI research field concerns *Usability*. Preece *et al.* (1994) describe it as "making systems easy to learn and easy to use". A more formal definition is proposed in ISO 9241-11 (ISO, 2005), as "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". This has

[1]Linked Data: http://linkeddata.org/
[2]DBpedia: http://dbpedia.org/
[3]Wikipedia: http://wikipedia.org/

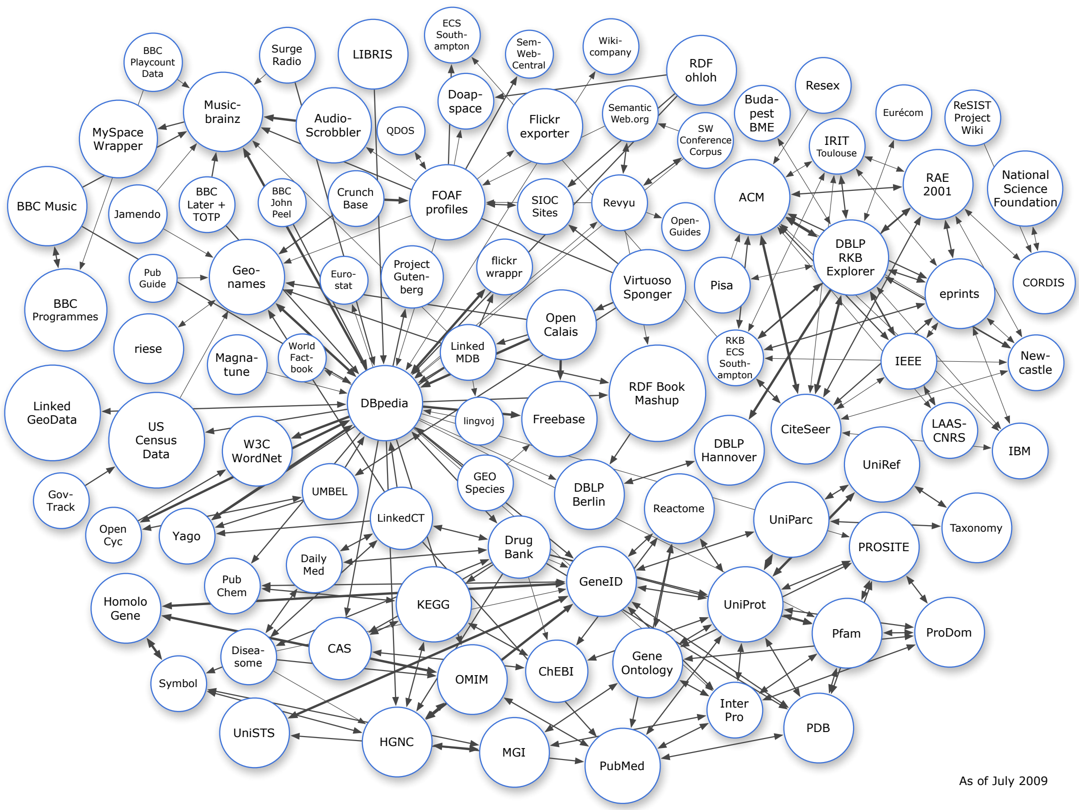


Figure 2.4: Linking Open Data Cloud

been studied from different perspectives such as engineering practices, testing, etc., in order to analyse users' dexterity and satisfaction when interacting with a computer system's user interface.

Hence, the adequacy of user interfaces to goals and tasks is of the uttermost importance for their successful completion. For this to happen, users are faced with software and hardware artifacts, from keyboards, mouses, and computer screens, to windows, drop-down boxes, and text input fields. The combination of the users' dexterity and knowledge on interacting in these scenarios defines the basic set of assumptions upon which the usability of a user interface can be studied and measured. When software artifacts get in the way of users' goals, the effectiveness of their tasks decreases. Ultimately, this results in user frustration and – more often than not – in giving up finishing the tasks.

A particularly significant user group where these issues arise in greater detail concerns those with disabilities (Barreto, 2008). While often perceived as a minority *in sensu lato*, over 15% of the world population has some kind of impairment. When users with disabilities are interacting with computers, they have the inherent requirement of overcoming physical, sensorial, or even cognitive disabilities imposed by the user interface. This is often performed with the aid of specialised hardware and/or software. Concerning the software side of this question, operating systems can provide facilities to level the interaction ground for all users. One such example is the use of a screen reader by users with visual impairments. Screen readers render speech equivalents of visual-centric interfaces (e.g., reading aloud textual content, informing about dialogue structure, etc.), and provide interaction input capabilities such as keyboard-based navigation.

However, developers and designers still have the onus of designing and implementing user interfaces that provide affordances for people with disabilities, i.e., that can be accessed and interacted by the users without diminishing their performance and effectiveness. The availability of these *online ramps*, thus, opens the way towards the *accessibility* of software applications and, ultimately, the improvement of user experience for people with disabilities.

In the context of information and communication technologies, the term *accessibility* is often understood as the *ability to access* resources, no matter the difficulties that might stand in the way of such tasks. It is a concept that is commonly associated to people with disabilities such as the blind, the deaf, etc. Lesser known audiences include people with cognitive disabilities (e.g., the reading impaired), the elderly – which typically suffer from several disabilities, albeit mildly –, amongst others.

Accessibility is not just a specific, closed-form software engineering field. It is also present in the scope of general software development, where providing access to everyone is pushed in different fronts. Standards such as ISO/TS 16071:2003 (ISO, 2003) and legislation such as USA's Section 508 (US Congress, 1998) open the way to enforce an *inclusive design* approach (Keates & Clarkson, 2003) at different levels: software applications, Web sites and applications, video and multimedia products, etc. These standards and laws conform to the underlaying principle of accessibility that the level of user experience must be the same for everyone, independently from impairments.

To cope with such range of audiences, a myriad of interaction devices and software capabilities might be used (both for input and output modalities) when interacting with a user interface. For instance, people with vision impairments benefit from using voice recognition and keyboards as input methods, and screen-reading as the preferred output. In the opposite way, people with hearing impairments benefit from pure visual outputs. But when considering the whole ecosystem of audiences, multiple combinations of inputs and outputs have to be taken into account when providing accessibility features to user interfaces. This way, a proper accessibility quality becomes more affordable.

While accessibility is mostly viewed as a way to end with implicit discriminations against people with disabilities on accessing information and working with software in general, these principles can also be applied in a broader scope (Harper *et al.*, 2006). For instance, the hardware limitations of mobile devices pose difficulties on interacting with a user interface, such as limited computational capabilities, or small display size. Another example concern situations where users are unable to interact visually with a user interface, such as while driving, running, etc. Thus, accessibility's benefits can also be expanded to a more universal – and literal – sense, i.e., the *ability to access*.

From the perspective of the Web, due to its hypertext nature, user interfaces – Web pages – must be created in such a way that anyone can perceive and navigate within their contents, e.g., the availability of alternative content that can be perceived by anyone with a sensorial or cognitive impairment. For instance, in the case of people with visual impairments, screen readers can capture the summary of a table, thus providing the same clues about that particular content to all users. Consequently, it is important to adequate the content, structure, navigation, and interaction capabilities of Web pages through *inclusive design* approaches.

2.3.1 Stakeholders

The above Section already detailed some important players within the accessibility realm: end-users and developers. However, depending on the application content within the creation of a service or application that is put on the Web, several stakeholders might have a saying in the way it is created in a more accessible way, including:

- *Developers* (not just software developers, but also designers and team managers) that have the role of creating accessible Web sites. Their knowledge on accessibility is often less than desired, for which accessibility verification tools can report insightful clues and repair tips to increase accessibility;

- *Accessibility Assessors*, i.e., experts on accessibility, who are involved in analysing the adequacy of existing Web sites in what respects to their accessibility quality;

- *Public Bodies/Governmental Agencies* who have the obligation, often by law, to provide online information and services that do not hinder nor discriminate users, including those with disabilities;

- *Service Providers*, e.g., public and private enterprises and organisations, who set the goal of reaching bigger audiences and increasing their market share. Since, as pointed out earlier, 15% of the world population suffers from some kind of impairment, accessibility can be perceived as a strength and opportunity that can distantiate themselves from their competitors;

- *Elderly and Disabled Users* who benefit from enhanced accessibility and are the actual potential end-users for every single Web site.

To doubt everything or to believe everything are two equally convenient truths; both dispense with the necessity of reflection.

Henri Poincaré

Chapter 3

State of the Art

3.1 Introduction

The previous Chapter has presented the core concepts of how the Web is architected, as well as how accessibility is woven into it. This Chapter discusses the most important work that has been done until know in the realm of Web Science and Web accessibility, in what regards to analysing the Web from a user perspective. Next, both of these themes are explored more deeply.

3.2 Web Science

Defined by Berners-Lee *et al.* (2006) as the *science of decentralised information systems*, the Web science discipline[1] strives for the study of the Web as a first-order organism (Hendler *et al.*, 2008). This often refers to the application of several software *lenses* for its analysis, i.e., how macroscopic properties, such as PageRank (Brin & Page, 1998), emerge from its basic microscopic properties. The key challenge lays at understanding how this poses changes on the way people perceive and experience the Web and, conversely, how people change the Web itself. For instance, in the case of PageRank, ranking properties emerge from the microscopic action of creating Web pages containing links between Web pages.

With this goal in mind, and its reflection into the realm of analysing the Web's accessibility quality, forthcoming texts present the the main contributions in Web science: its *lifecycle*, its most relevant *characterisations*, and what is the role of *linked data* in the research field.

3.2.1 The Web Science Lifecycle

In order to attain the goals of Web science, the Web must be studied from a multitude of research fields, often in a inter-disciplinary way (Shadbolt, 2008). This includes studying its technical properties, e.g., trough computational and mathematical perspectives, along the side of their impact on society, e.g., social, economic, and legal perspectives. This intertwine between both factors, coupled

[1] Web Science Trust: http://webscience.org/

with the complexity inherent of the size of the Web, is reflected in the Web science lifecycle, as depicted on Figure 3.1.

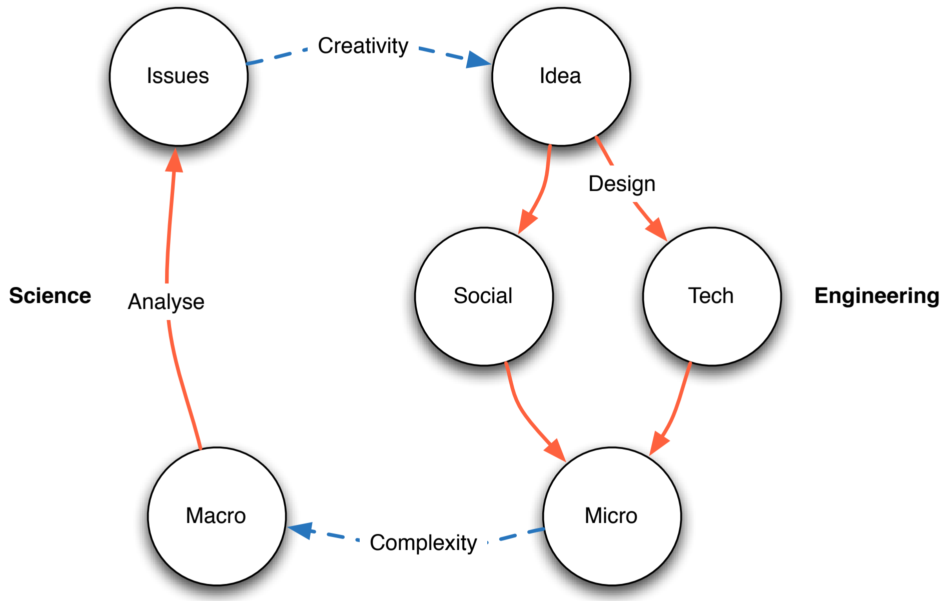


Figure 3.1: The Web Science Lifecycle (adapted from Berners-Lee (2007))

This lifecycle defines the constant evolution of the Web, from its research perspective: *Ideas* pave the way to *Social* conventions and *Technological* solutions, which are directly reflected at a *Microscopic* level (e.g., Web site); these factors scale towards a *Macroscopic* level as emergent phenomena, due to its *complexity* properties (i.e., size and interconnectedness); in turn, this jump in scale brings new analysis challenges, which raise a new set of *Issues* concerning the Web; lastly, fuelled by *creativity*, these issues are turned into new ideas, thus completing the Web science cycle. *PageRank*, as briefly explained above, is one of such cases: the *link-based ranking* emerges as a macroscopic property of the microscopic linking activity; this linking activity is seen as both a technical feature (the link itself, inherently provided by HTML and the Web architecture), and its social counterpoint, i.e., the *meaning* of the link (e.g., referencing information stated on another Web page).

This high-level framework is verified in several real cases of the Web, such as Wikis, search engines, spam, etc., as detailed in Berners-Lee (2007). Hence, human intervention has been observed as directly influencing the Web's structure and content, as well as the opposing way.

3.2.2 Web Characterisation

Several researchers have also tackled the analysis of the Web from a more analytical approach, in order to study the way the Web influences users, and vice-versa. The Web can be perceived as a graph of interconnected Web pages, where hyperlinks serve as directed edges. With its sheer size in the order of billions of Web pages, it is important to analyse its topology, since it can shed a light on how people create and use the Web. PageRank, as presented previously, is one of the key examples on how understanding the Web's topology leads to emergent behaviour from the Web, which is subsequently exploited by modern search engines.

Dill *et al.* (2002) have shown that the Web follows a fractal nature (as presented in Figure 3.2): it is composed by a strongly connected cluster (SCC) of Web pages and two other portions which it links to (OUT) and is linked from (IN); analysing SCC, this pattern is found once again, with several SCCs having the same shape.

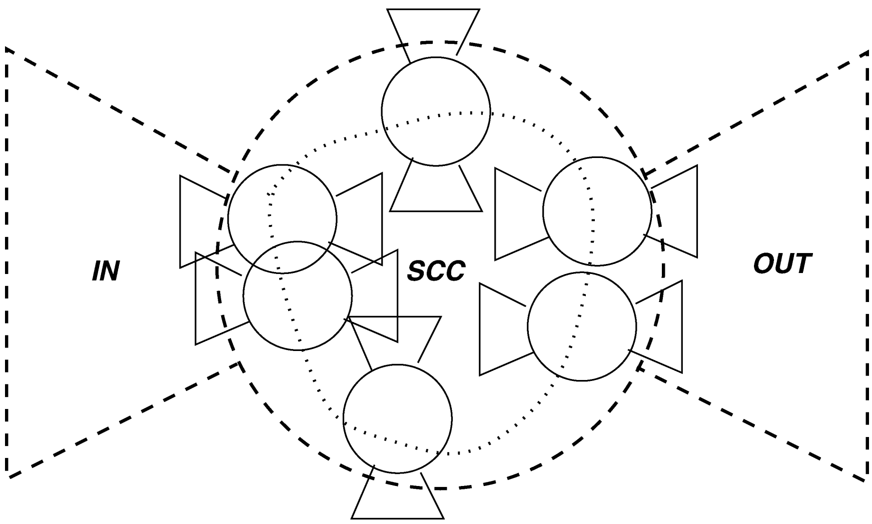


Figure 3.2: The Fractal Nature of the Web

Interesting portions of the Web that maintain this fractal property of the Web concerns national Webs (Baeza-Yates *et al.*, 2007). For instance, the outdegree of a Web page – how many hyperlinks are contained within a Web page – is a constant property for different national Webs, i.e., have the same linking structure. As depicted in Figure 3.3, the distribution of links follows a power law for all national Webs (few Web pages have many links, whereas many Web pages have few links).

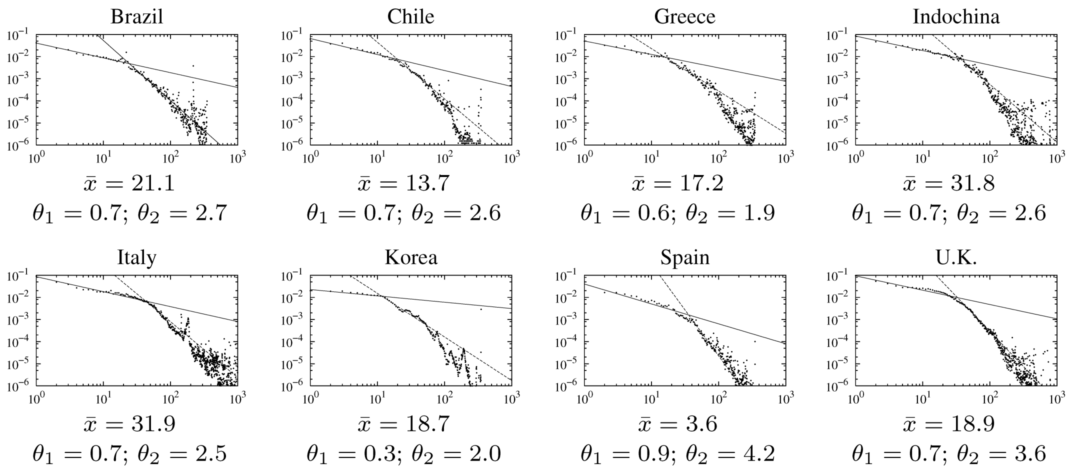


Figure 3.3: Similarity of Outdegrees Distribution in National Webs (X axis represents page count, Y axis represent link count)

This aggregation-based analysis of Web properties can also shed a light in domains other than hyperlinking. The inherent content characteristics of each Web page (e.g., visual structure, coherence of navigation, etc.), as well as of user navigation behaviours, are both factors that deeply influence how people perceive and interact with the Web. For instance, a long-term study by Weinreich *et al.* (2006) has shown the average stay times on Web pages is very low, where more than 50% of the Web pages had a stay time period of less than 12 seconds (Figure 3.4).

This study has also shown a direct correlation between this time and hyperlink activity areas, where the area of Web pages that had more interactions (i.e., *clicks*) are located in the *top left* corner of Web browsers – the most immediate content location perceived by users (Figure 3.5).

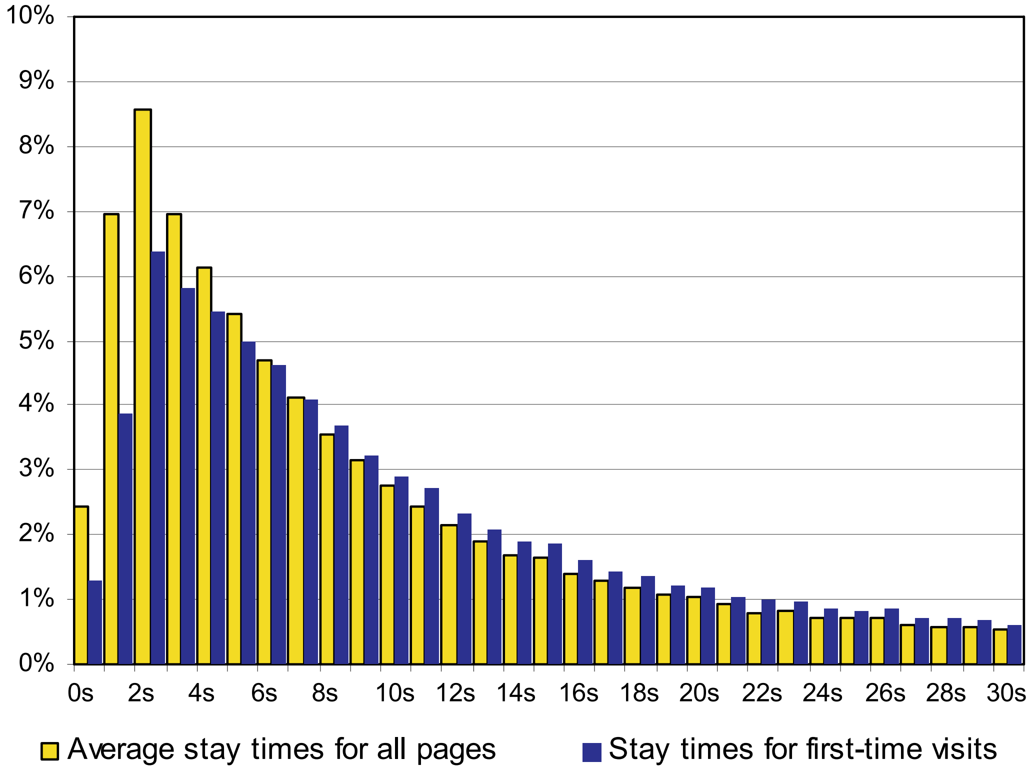


Figure 3.4: Distribution of Stay Time by Users

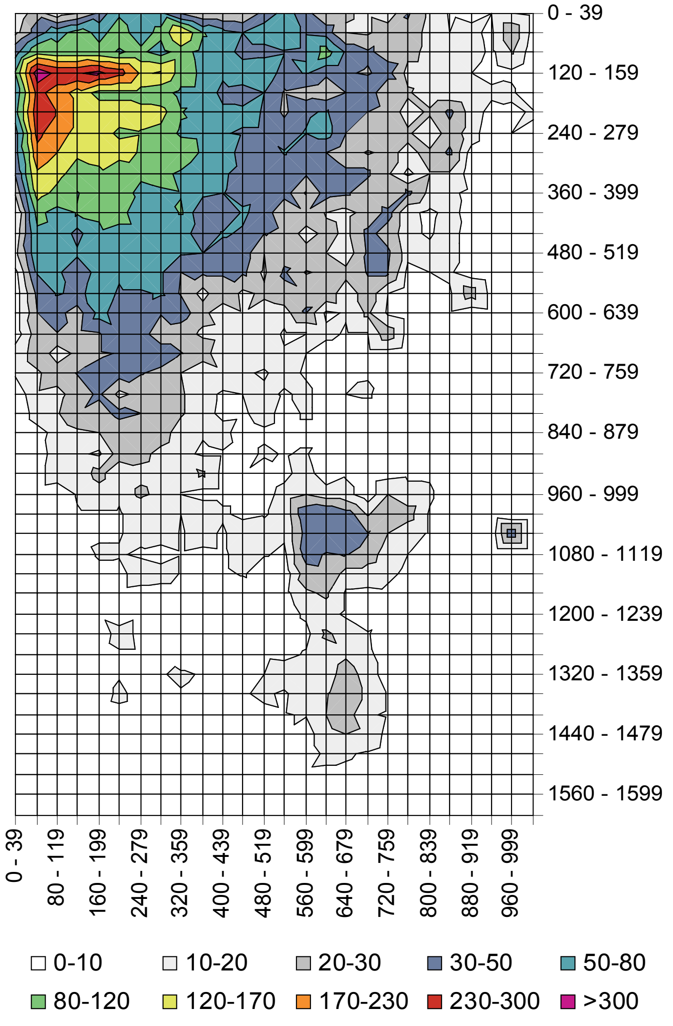


Figure 3.5: Link Activity Areas of Web Users

3.2.3 Linked Data

As described in the previous Chapter, *Linked Data* is starting to play a strong role on interconnecting data sources around the Web. By leveraging the Web's architecture and Semantic Web technologies, it helps create a common, strongly interrelated *Web of data*. Figure 3.6 shows how Linked Data fits into the Web science lifecycle. The ever increasing availability of data in the form of Semantic Web technologies (e.g., RDF, OWL), coupled with the problem of not having a homogenous availability of these data as first class citizens on the Web lead to the Semantic Web's reframe as Linked Data. Consequently, it is expected that *Linked Data* forms a solid ground to a data corpus that can be analysed macroscopically.

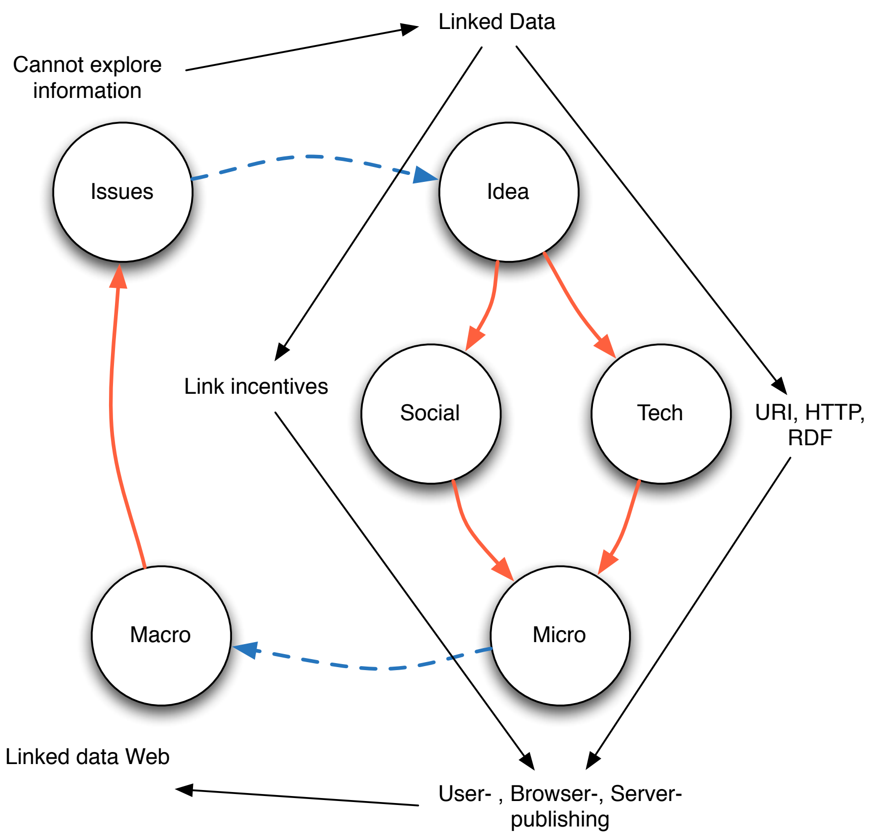


Figure 3.6: Link Data in the Web Science Lifecycle

This reframing of Linked Data into the Web Science lifecycle, as explained, is based on defining data linking incentives (e.g., data enrichment from disparate sources) that can be exploited through its publication and availability directly on the Web. This approach to data sharing and linking paves the way towards more structured macroscopic analysis of the Web. A study over Wikipedia has shown that semi-structured information can be leveraged into *Linked data* concepts, which resulted on DBpedia (Bizer *et al.*, 2009). The availability of this data set as *Linked data* has put it as the dominant source, where 33% of RDF triples are connected to it (Halpin, 2009). Having this data available is starting to foster the creation of applications built on top of the linked data set, from interactivity such as exploratory user interfaces (Figure 3.7), to more analytical approaches such as geographic terms disambiguation (Cardoso & Silva, 2010).

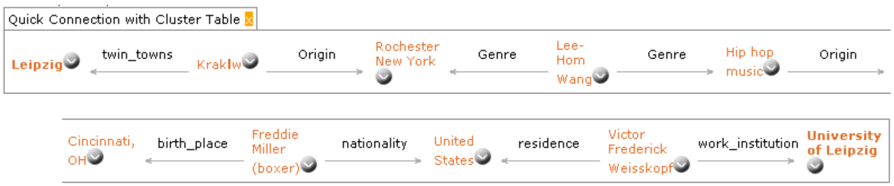


Figure 3.7: Exploring Concept Relationships on Wikipedia

3.3 Web Accessibility

Web Accessibility emerges as both a natural extension and specialisation of applying general accessibility concerns to the Web browser and Web metaphors, as presented in the previous Chapter. Since the Web has its own idiosyncrasies in what respects to multimedia content production and availability, hyperlinking, and its architecture, it is important to understand how its building blocks are adequate towards people with disabilities.

3.3.1 Standards and Guidelines

The Web accessibility discipline focuses on providing an accessible path to Web content production and consumption in par with the user experience typically found for non-impaired users. The WAI model (Henry, 2005) for this vision is centred on framing accessibility technologies between three main components, as presented in Figure 3.8: the *Developer*, the *User*, and Web *Content* (e.g., Web pages, media, etc.)

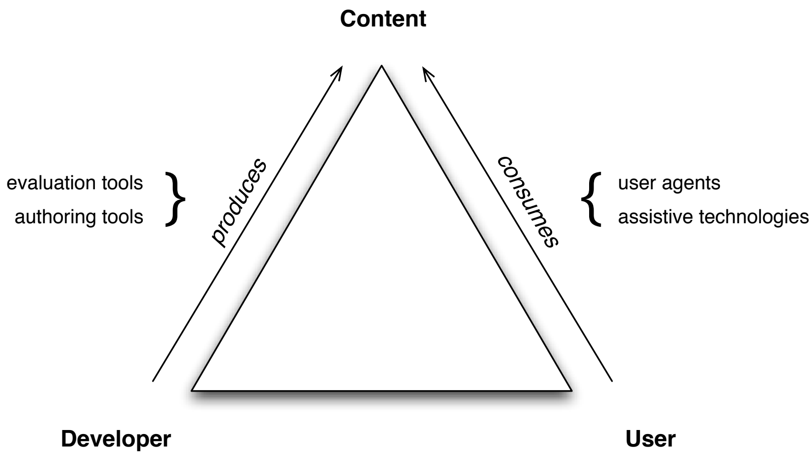


Figure 3.8: Components of the WAI Model

Each interaction within the WAI model has a corresponding set of well-defined guidelines that, if properly applied, improve the experience of users with disabilities on the Web. The depicted interactions are detailed as follows:

- Developers *produce content* that must be accessible. For this, they must conform with the Web Content Accessibility Guidelines (WCAG) (Caldwell *et al.*, 2008), a process that is typically aided by *evaluation tools*. Furthermore, this can also be supported by *authoring tools* that produce accessible content, as defined by the Authoring Tools Accessibility Guidelines (ATAG) (Treviranus *et al.*, 2009);

- Users *consume* content within Web pages through *user agents*, such as Web browsers and media players, with the aid of *assistive technologies*, e.g., screen readers or braille displays. Consequently, these entry points for the Web must also be accessible, i.e., conforming with the User Agent Accessibility Guidelines (UAAG) (Ford *et al.*, 2009).

In essence, all of these guidelines from the WAI – WCAG, ATAG, and UAAG – concern a different aspect regarding the adequacy of a Web page to users' impairments. Either directly or indirectly, all interactions within the WAI model are dependent on the accessibility quality of a Web page. Next, each one of these guidelines are discussed more thoroughly.

3.3.1.1 WCAG

WCAG, the Web Content Accessibility Guidelines, serves as the basis to verify the accessibility compliance that a Web page must have, i.e., a *quality mark*, in order to provide an adequate experience for users with disabilities. WCAG is mainly targeted at designers and developers responsible for Web front-ends – the *visible* side of the Web –, by listing a set of guidelines for accessible *ramps* in Web content. An example is WCAG 1.0 Guideline 1, which states:

> *Provide equivalent alternatives to auditory and visual content.*

Each guideline is further decomposed into *checkpoints* that assess a different aspect of the guideline. Based on the example above, an excerpt of Checkpoint 1.1 states:

> *Provide a text equivalent for every non-text element (e.g., via `alt`, `longdesc`, or in element content).*

In order to translate each checkpoint into concrete support for Web content, WCAG defines a set of techniques that can be applied to different Web technologies, such as HTML or CSS (Chisholm *et al.*, 2000). An example technique for Checkpoint 1.1 is presented below, in the form of an HTML excerpt.

```
<A href="routes.html">
  <IMG src="routes.jpg" alt="Current␣routes␣at␣Boulders␣Climbing␣Gym">
</A>
```

Listing 3.1: An HTML Technique for WCAG 1.0 Checkpoint 1.1

Here, the technique presents the `alt` attribute of the HTML `IMG` element, which affords an *alternative textual* description of what is conveyed in the linked image. By following WCAG techniques, designers and developers have a concrete, expert-created guide to provide barrier-free Web pages to users with disabilities.

Each checkpoint in WCAG is also mapped into one of three priorities, depending on the severity introduced by the respective non-compliance, as follows (quoting from WCAG 1.0):

- *Priority 1.* A Web content developer must satisfy this checkpoint. Otherwise, one or more groups will find it impossible to access information in the document. Satisfying this checkpoint is a basic requirement for some groups to be able to use Web documents.

- *Priority 2.* A Web content developer should satisfy this checkpoint. Otherwise, one or more groups will find it difficult to access information in the document. Satisfying this checkpoint will remove significant barriers to accessing Web documents.

- *Priority 3.* A Web content developer may address this checkpoint. Otherwise, one or more groups will find it somewhat difficult to access information in the document. Satisfying this checkpoint will improve access to Web documents.

More recently, WCAG 2.0 (Caldwell *et al.*, 2008) has been superseding the 1.0 version, where it reviews the evolution of Web technologies in the light of their accessibility issues and technological support. WCAG 2.0 also introduces the notion of *principles*, which aggregate guidelines into four *POUR* categories: *Perceivable*, *Operable*, *Understandable*, and *Robust*.

3.3.1.2 ATAG

ATAG, the Authoring Tools Accessibility Guidelines, are focused on the creation of accessible Web content, as well as on the availability of content production platforms that can be used by people with disabilities. It is targeted to the backends of Content Management Systems such as Wordpress[1] and Mediawiki[2],

[1]Wordpress: http://wordpress.org/
[2]Mediawiki: http://www.mediawiki.org/wiki/MediaWiki

as well as to Web development applications such as Adobe Dreamweaver[1] and Microsoft Expression Web[2].

Just like in WCAG, ATAG defines a set of *Guidelines*, which are decomposed into *Checkpoints* and *Techniques*. For instance, to afford the compliance with WCAG Checkpoint 1.1 (presented above), ATAG defines its checkpoint 3.1 as:

> *Prompt the author to provide equivalent alternative information (e.g., captions, auditory descriptions, and collated text transcripts for video).*

Based on this, ATAG also defines several techniques that authoring tools must provide to Web content producers, depending on the accessibility barrier tackled by the corresponding WCAG Checkpoints. For the example presented above, one of techniques states:

> ***Short Text Labels (for alternate text, titles, etc.):*** *These types of alternative equivalents require only short text strings from the author, so the prompts for them may be best located as text boxes within property dialogs, etc. An important consideration is that the function of the object (decorative, button, spacer, etc.) will be important to the instructions given to the author on what to write. The object function may be prompted for or discovered by automated heuristics.*

Also, due to the relationship with WCAG, ATAG also define three priority levels for checkpoints. These levels are directly related to satisfying content production according to the corresponding WCAG priority. For instance, an authoring tool can only be compliant with ATAG priority 2 if the content it produces is compliant with WCAG priority 2.

3.3.1.3 UAAG

UAAG, the User Agent Accessibility Guidelines, provides a description of techniques for a barrier-free interaction with Web content for users with disabilities.

[1] Adobe Dreamweaver: http://www.adobe.com/products/dreamweaver/

[2] Microsoft Expression Web: http://www.microsoft.com/expression/products/Web_Overview.aspx

While WCAG and ATAG are targeted at Web content production, UUAG is targeted at developers of Web browsers, assistive technologies, and other User Agents commonly used by people with disabilities. While it is of the uttermost importance for Web content to be accessible, if users have barriers to this content that are imposed by User Agents, the inherent accessibility purpose is defeated. As an example, UUAG Guideline 1 states:

> *Support input and output device-independence.*

By following this guideline, through its corresponding checkpoints (*Full keyboard access*, *Activate event handlers*, and *Provide text messages*), developers can ensure that accessible Web content can be used appropriately by users with disabilities without breaking their preferred interaction model.

3.3.1.4 Other Accessibility Standards

While the Web Accessibility Initiative, via the World Wide Web Consortium, is the most well known entry point upon which Web accessibility is framed, designers and developers might follow other standards, processes, or accessibility guidelines that are taken into account by accessibility stakeholders throughout the world, including:

- *Section 508.* This amend to the Rehabilitation Act of 1973 puts a strong emphasis on making Federal information and services accessible to all USA citizens. A particular technical standard in *Section 508* concerns "Web-based Intranet and Internet Information and Applications", where a set of guidelines have been devised how to meet the requirements for accessible Web sites and applications (similarly to WCAG). More recent versions of Section 508 are currently being reviewed according to WCAG 2.0 (e.g., application scenarios for WCAG 2.0 techniques);

- *BITV.* Also known as *Barrierefreie Informationstechnik-Verordnung*, BITV is the official standard for barrier-free information access in Germany. It states a set of *Standards* and *Requirements* for accessibility compliance that are based on WCAG 1.0, but applied to a broader scope (i.e., includes non-Web content);

- *Resolução do Conselho de Ministros nº 155/2007.* The Portuguese governmental directive for Web Accessibility, which states that all governmental and public bodies Web sites must comply with WCAG 1.0, levels A or AA (depending on the existence of transactional processes).

Apart from Section 508, which has a broader scope than the Web, most national laws on information accessibility rely on WCAG (mostly its 1.0 incarnation) and associated guidelines as the technical basis for accessibility adequacy of public services' Web sites[1].

3.3.2 Evaluation

There are three principal ways to understand this adequacy, by means of evaluation procedures: *user studies*, *expert analysis*, and *software evaluation*, as shown next in Figure 3.9.

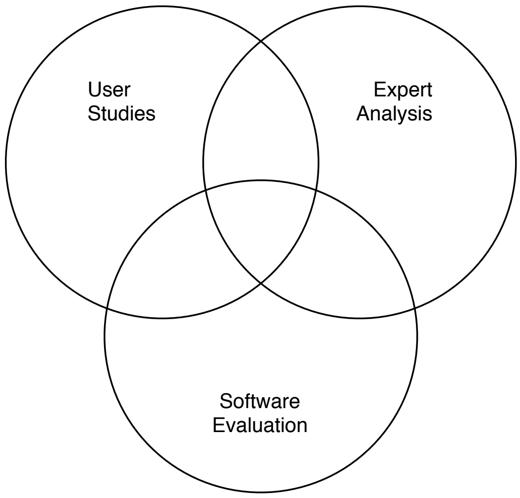


Figure 3.9: The Three Evaluation Methods of Web Accessibility Adequacy

The WAI process details when and how each evaluation procedure should be used, weighting the pros and cons of each evaluation procedure. The most thorough process is *user studies*, where a mix of techniques such as usability studies, interviews, ethnography, etc. help reaching an overall, in-depth vision

[1]Policies Relating to Web Accessibility: http://www.w3.org/WAI/Policy/

of the accessibility quality of a given Web page. While *user studies* can provide the biggest insights on the accessibility of a Web site, the cost, complexity, and expertise required for such studies forbids most designers and developers to assess Web sites.

Consequently, WCAG was devised also as an expedite way to evaluate Web sites from their accessibility perspective. Based on WCAG's guidelines, with the aid of WCAG evaluation *techniques*, designers and developers can evaluate a Web page's accessibility quality and have hints on how to fix its accessibility barriers.

3.3.2.1 The Expertise Factor

WCAG is based on the assumption that *expert analysis* is performed by developers or designers, in order to meet the requirements stated in each checkpoint. Therefore, the effort and accuracy of designing accessible Web pages is directly related to the level of expertise of designers and developers. If they cannot assess correctly the potential accessibility problems, they will create Web pages with decreased accessibility quality. However, recent studies have shown that:

- Deep knowledge on accessibility issues and their reflection on Web front-end technologies – *high expertise* – is paramount for the success of verification and application of WCAG techniques, as shown by a meta-evaluation study comprising 19 experts and 51 non-experts on Web accessibility (Yesilada *et al.*, 2009);

- For non-experts, the reliability of verification is less than desired, especially when subjectiveness on the evaluation process is implied (Brajnik, 2009);

- Both experts and non-experts produce too many false positives and miss actual problems of accessibility in Web page analysis, which results on ambiguity for human inspection processes (Brajnik *et al.*, 2010).

3.3.2.2 Evaluation Tools

Implementing Web accessibility evaluation procedures based on the techniques defined in WCAG is not a trivial task. Each Web page is a mix of different technologies, such as HTML, CSS, and Javascript, which blend together into its final representation within a Web browser. Thus, to evaluate a Web page, it is (ideally) necessary to inspect and interpret how it contains accessibility barriers. Some of these evaluation techniques are objectively implementable, e.g., as detailed in UWEM (Velleman *et al.*, 2007) with XPath expressions (Listing 3.2):

```
//img[@alt='']
//area[@alt='']
//input[@type='image'][@alt='']
//applet[@alt=''][count(local-name(*)!='param')=0]
//object[count(local-name(*)!='param')=0]
```

Listing 3.2: UWEM XPath Verification for WCAG 1.0 Checkpoint 1.1

However, there are cases where evaluation still depends on human judgement, such as matching an image with its textual description. To make the distinction between fully automateable techniques, these tools often make the distinction between *errors* and *warnings* in the techniques' application outcome. *Errors* occur when a given tool is certain that a given accessibility problem has been found within a Web page (e.g., the lack of an `alt` attribute in an `img` element). *Warnings* are raised when there is a probability of having an accessibility problem, which has to be further verified by the developer or designer.

This disparity can lead to completely different interpretations on whether a Web page meets a specific accessibility criterion. The implementation of automated software analysis of Web accessibility paves the way towards a more strict and objective analysis (all criteria are applied in the same way, independently of each Web page), but at the cost of not being capable of fully analysing the accessibility outcome for several cases. Nevertheless, all these software tools can be leveraged by developers and designers in order to guide them in the creation of more accessible Web pages, e.g., according to WCAG guidelines and corresponding techniques.

The WAI provides a thorough compilation of Web accessibility evaluation software tools available[1]. The list includes a miscellany of software comprised of a disparate set of features, goals, and usage.

[1] Web Accessibility Evaluation Tools: http://www.w3.org/WAI/ER/tools/complete

We analysed the 127 tools on this list by focusing on some of the technological dimensions that directly influence the selection of a software for Web accessibility evaluation: *automation*, *crawling*, *technologies*, *guidelines*, *reporting*, *server capabilities*, and *licensing*. Next, we detail how each one of these dimensions is important in the overall context of software-based evaluation, as well as how the analysed tools fit into each dimension.

Automation Automating the evaluation of accessibility adequacy of Web pages is a key feature for scaling these practices. More than half of the tools (nearly 57%) provide some kind of automated evaluation capability. Non-automation -centric tools include disability simulators (e.g., daltonism, screen blurring, etc.) From automation-capable tools, only 21% offer some sort of crawling feature, to facilitate the evaluation of large sets of Web pages.

Technologies Another dimension concerns the client-side technologies and media used to create Web pages. Figure 3.10 presents the top technologies that are analysed by evaluation tools.

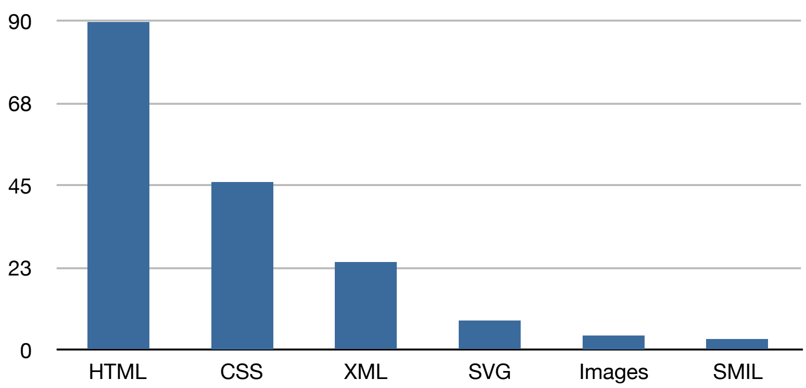


Figure 3.10: Distribution of Technologies Analysed by Evaluation Tools

We discovered that the trend of this distribution closely resembles an exponential decay. The clear winner in technologies is, as expected, HTML, with 90 tools (71%) implementing some kind of analysis of Web pages' structure and content, showing the tendency on implementing evaluators mostly focused on HTML.

Guidelines Independently of technology targeting, evaluation tools typically centre around a set of guidelines for accessibility compliance. In Figure 3.11 we present the distribution of the implemented guidelines.

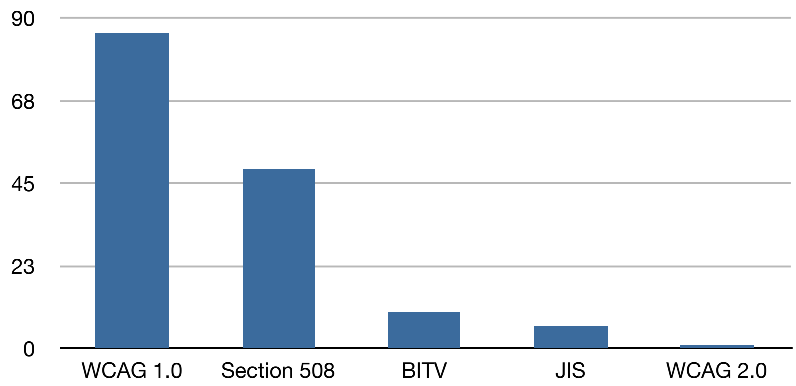


Figure 3.11: Distribution of Guidelines Implemented by Evaluation Tools

As expected, WCAG 1.0 leads the way on guideline implementation by evaluation tools (68%). The main reason for this fact is twofold: first, several Web accessibility laws around the world are based on WCAG 1.0 (W3C, 2006); and second, WCAG 2.0 has only recently reached *Technical Recommendation* status at W3C. It is worth mentioning also that these numbers are limited for WCAG 2.0 due to an incompleteness of the list provided by the WAI. Country-specific guidelines are also implemented by some tools, where United States' Section 508 leads the way (39%).

Reporting Another factor for Web accessibility evaluations concerns the interpretation of evaluation results. While human-readable formats are important for the development of accessible Web pages, machine-friendly formats allow for wider inspection of Web accessibility properties at the large. Figure 3.12 depicts the corresponding data about reporting capabilities.

Out of the 127 tools analysed, 80 of them provide some kind of reporting capability (63%), whereas the rest of the tools are centred either on document format conversion (e.g., Web page transcoding (Lunn *et al.*, 2008)) or on disability simulation. Out of these, few support machine-friendly formats: only 14 (11%) provide compatibility with the EARL format (Abou-Zahra & Squillace, 2009), while 7 (6%) with comma separated values (CSV).

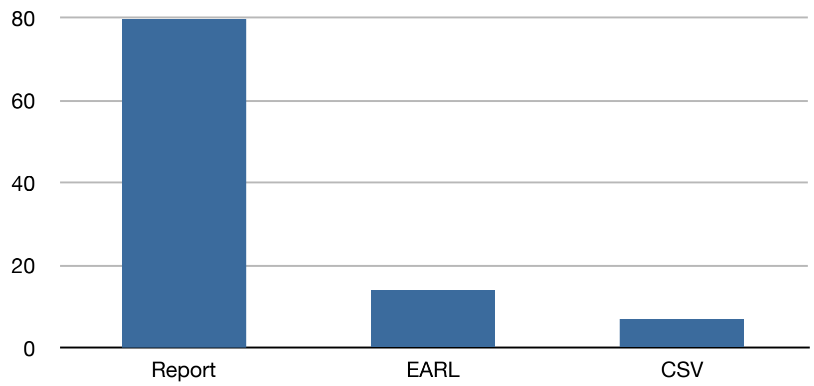


Figure 3.12: Distribution of Reporting Capabilities by Evaluation Tools

Server Capabilities Performing large-scale, software-based evaluations of accessibility on the Web, requires a high computation capacity, since it involves the analysis of millions of Web pages. Consequently, these tasks are often delegated to servers and computing clusters. In this scenario, we also analysed how tools are adequate to being deployed into servers, as presented in Figure 3.13.

Figure 3.13: Distribution of Server Capabilities by Evaluation Tools

We verify that only 51 tools are available online with a Web interface (40%). Out of these, 14 are available as a downloadable server (11%), and just 2 provide some kind of programmatic access through a Web service interface (2%).

Licensing Lastly, the final dimension analysed comprises the chosen software licensing model. We made the assumption that one of the factors that afford tool-based evaluation of Web accessibility concerns the availability of evaluation tools. Besides, open source licensing is also a key factor for the technological development of evaluators. Figure 3.14 depicts these properties.

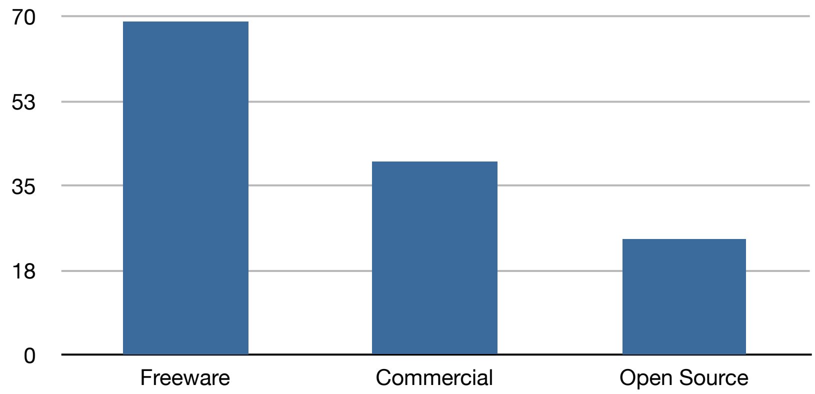


Figure 3.14: Distribution of Software Licensing by Evaluation Tools

The *freeware* licensing model is the preferred option for 69 evaluation tools (54%), whereas its opposite, *commercial* licensing is present at 40 evaluation tools (32%). The availability of the source code for evaluation tools is a key feature for 24 of them (19%).

3.3.2.3 Metrics

To ensure the compliance with WCAG, the WAI has devised a quality check mark comprised of three levels – A, AA, and AAA – which directly map onto WCAG priorities, as follows:

- *Conformance Level A*: all Priority 1 checkpoints are satisfied;
- *Conformance Level Double-A*: all Priority 1 and 2 checkpoints are satisfied;
- *Conformance Level Triple-A*: all Priority 1, 2, and 3 checkpoints are satisfied.

While this simple metric affords the distinction of the accessibility quality between two different Web pages, its coarseness does not make distinction between a Web page that fails a single WCAG priority 1 checkpoint once, and another Web page that fails to comply with all priority 1 checkpoints. While this property of accessibility quality assessment is useful for procurement and law abidance reasons, to better understand the accessibility quality of Web pages, there is the need to properly quantify accessibility in a more detailed way. With this goal in mind, several Web accessibility metrics have been devised by several researchers. Below, we discuss the most relevant ones: *Failure Rate*, *WAB*, *UWEM*, *A3*, and *WAQM*.

Failure Rate. Sullivan & Matson (2000) proposed *failure rate*, one of the first quantification metrics for Web accessibility. This metric allows for understanding the weight of errors (i.e., *failures* in its terminology) from the perspective of *potential failure points* (i.e., HTML elements that could allow for accessibility barriers), i.e.:

$$Failure\ rate = \frac{Actual failures}{Potential failures}$$

The outcome from the quantification that this metric provides is rate between 0 and 1 (i.e., percentage) that states the inadequacy of accessibility for a given Web page. The authors implemented 8 WCAG 1.0 checkpoints (priority 1) to assess a set of Web pages and analyse the outcome of their metric. This way, a Web page failing checkpoint 1.1 5 times in 5 points is less accessible – and has a higher failure rate – than other page failing 4 times in 5 points.

WAB. WAB (Zeng, 2004) (Web Accessibility Barrier) builds upon Failure rate, but introduces the notion of weighting the priority of checkpoints:

$$WABScore = \sum_{b=1}^{n} \frac{Actual failures}{Potential failures \times priority}$$

This metric aggregates the rates for each barrier b (which represents a checkpoint), which results in a quantification with no upper bound.

UWEM. UWEM (Velleman *et al.*, 2007) (Unified Web Evaluation Methodology) is another metric for quantifying the accessibility quality of Web pages that leverages weighting functions to cope with severity levels:

$$UWEMScore = \frac{1}{T}\sum_{j=1}^{T} 1 - \prod_{b}(1 - \frac{n_{bj}}{N_{bj}W_b})$$

Opposing to Failure rate, UWEM returns a rate between 0 and 1, where *1* represents the highest accessibility quality that can be measured, i.e., passing all accessibility checks.

A3. A3 (Bühler *et al.*, 2006) has been devised as an extension to UWEM that can cope with the specificities of user-tailored accessibility evaluation:

$$A(u) = 1 - \prod_{b}(1 - R_b S_{ub})^{C_b}$$

Where C_b represents the notion of rates for each accessibility barrier that is being verified. The resulting outcome, $A(u)$ is also a rate between 0 and 1, but specific for a user u (or, more generally, to a set of users represented by the potential barriers that can limit their access to information).

WAQM. WAQM (Vigo *et al.*, 2007) (Web Accessibility Quantitative Metric) provides a more thorough, detailed analysis on the effects of quantifying the accessibility level of a given Web page. This metric is calculated through an algorithm that aggregates and weights different aspects: the *POUR* principles of WCAG 2.0 (translated into WCAG 1.0 checkpoints), the type of analysis for each checkpoint (*error* or *warning*, i.e., its capability of being fully automated), as well as priority levels. The WAQM metric also assumes an equivalency of failure ratio between errors and warnings, assuming that the rate of errors that cannot be checked automatically is similar to those of detected errors.

3.3.2.4 EARL

Evaluating Web accessibility is a process that requires inspection of evaluation results. More often than not, as detailed earlier in Figure 3.12, these results are presented as (or close to) free text towards developers and designers' interpretation and inspection. While this is a very valid proposition, it provides severe obstacles to machine-based analysis.

To mitigate this problem, and based on the importance and multiple utility that is dependent on compliance checking with regards to WCAG, the Web Accessibility Initiative is currently defining a standard way to describe this compliance: EARL, Evaluation and Report Language (Abou-Zahra & Squillace, 2009). This language allows developers, end users, accessibility advocates, etc. to have a common language (which is also machine-processable) to support the aforementioned uses cases, as detailed in the EARL Guide (Velasco & Koch, 2009).

EARL is a dialect of concepts built on top of RDF, that centres on describing general evaluation results as *Assertions* composed by: *Assertor* (who/what evaluates), *Test Subject* (what is being evaluated), *Test Criterion* (which evaluation), and *Test Result* (the evaluation's outcome). An example EARL assertion is depicted below in Listing 3.3:

```
<earl:Assertion rdf:about="#assertion">
  <earl:assertedBy rdf:resource="#assertor"/>
  <earl:subject rdf:resource="http://www.example.org/"/>
  <earl:test rdf:resource="http://www.w3.org/TR/WCAG20-TECHS/H36"/>
  <earl:result rdf:resource="#result"/>
</earl:Assertion>
```

Listing 3.3: An example EARL assertion (in RDF/XML serialisation format)

3.3.2.5 Sampling Methods for Web Site Accessibility Evaluation

Most of the work conducted on evaluating the accessibility of Web pages is performed manually. Even when (semi-)automated evaluation tools are used, the computational cost – in terms of time and resources – required for larger pools of Web pages such as a Web site typically becomes high. Thus, when aiming at Web site -level evaluations of accessibility, it can be impractical or even impossible to evaluate the accessibility quality of every single Web page. To mitigate this issue, some researchers have been devising sampling methods that afford the selection of representative Web pages from Web sites, such as those defined in UWEM (Bühler *et al.*, 2006).

3.4 Discussion

This Chapter has presented different (and complementary) approaches that heavily influence the large-scale evaluation of Web front-ends with regards to their accessibility quality. Both *Web science* and *Accessibility* sides of evaluation and assessment have been discussed and envisioned as attainable goals for a Web that is more accessible and more supportive for people with disabilities. In the light of this goal, it is worth revisiting the main hypothesis raised in this document, and discuss how these research areas tackle it:

> ***Hypothesis:*** *Web accessibility evaluation at large scale will leverage new knowledge about the Web itself, its development and adequate mechanisms to raise proper awareness of designers and developers towards accessibility.*

The highly increasing expansion of the Web into new domains requires novel approaches on evaluating Web front-ends that can be studied at large-scale. However, as detailed in the previous Chapter, the *Web science* and *Web accessibility* research fields are still fairly disconnected. On the former, there is novel and pioneering work that focuses on understanding the impact of users on the Web, as well as the opposite. However, none has been found about the Web's accessibility properties. On the latter, the advancements on Web accessibility evaluation – particularly on its automateable part – point some directions on how Web accessibility is shaped. But this research is still lacking understanding if, at macroscopic levels, there are patterns that emerge and/or if predictions of accessibility quality are maintained (particularly due to focus solely on the evaluation of small, local samples).

While all of the research areas provide evidences towards this hypothesis, it is yet to be confirmed at the scale of the Web. This gap between the state of the art and the hypothesis is revisited below and further discussed in the light of the problems raised earlier in Chapter 1, as well as based on the individual discussions present at the end of each Section.

Current *Web accessibility* evaluation research focuses on understanding how different evaluation methods behave (e.g., metrics, sampling, etc.) Thus, they are applied to a fixed – and limited – sample of Web pages for benchmarking purposes. This locality does not afford the analysis, discovery, and comprehension of potential Web accessibility phenomena. On the other hand, *Web science* does provide clues that support this hypothesis, but outside the realm of accessibility.

We have discussed the role of the *expertise factor* in the context of Web accessibility evaluation, where several studies have shown that designers and developers typically do not have the *expertise* to verify and understand if several inaccessibility situations are present. However, these studies do not put the emphasis on understanding how this difference is impacted at large scale. Furthermore, existing Web accessibility metrics mix *failures* and *warnings* such that they hinder the perception of accessibility quality.

As presented earlier in the Chapter 2 (*Background*), accessible Web sites and applications are a focus of Web engineering practices, where classical modeling frameworks can be leveraged to create usable and accessible Web sites. However, little is known about its impact in a more general sense of the Web, i.e., whether users, content producers, and even developers and designers are aware of the usage of template mechanisms in order to deliver accessible Web sites.

Web science practices, such as Web metrics and Web topology, have shown that the Web follows very specific laws in terms of its structure. However, from what could be found in the state of the art, there has been no studies on how this influences accessibility.

3.5 Summary

This Chapter presented the *State of the Art* in what respects to the analysis of the Web from two perspectives: its characterisation and its adequacy to people with disabilities. There are several research results that have shown that (a) the Web has microscopic characteristics that have an impact on its macroscopic side, and (b) Web accessibility is still a nascent practice, especially in what respects to its evaluation side, and that stakeholders' unawareness is still the biggest problem facing the notion of a *Web for All*.

Consequently, in order to study the validity of the research hypothesis raised in this thesis, there is a need to bridge both sides of the equation, as depicted below in Figure 3.15: *Web Science* and *Web Accessibility*. This means the application of Web Science concepts and scales to Web accessibility, and introducing Web accessibility measurement and characterisation on Web Science practices.

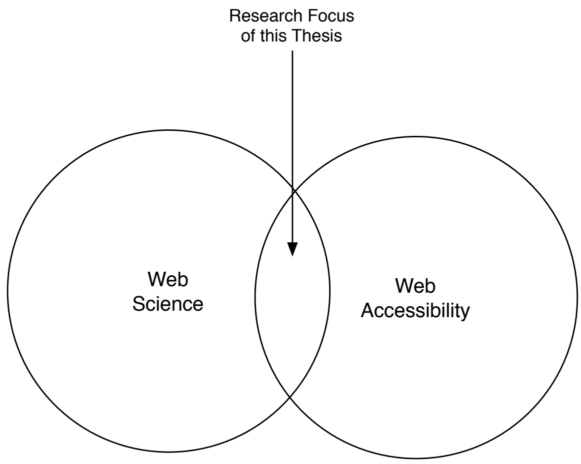


Figure 3.15: The Research Focus of this Thesis

The next Chapters detail the two main contributions of this thesis: (1) a set of experimental studies of Web accessibility analysis that afforded the discovery of some properties of the Web in what respects to its accessibility quality, and (2) how to model, implement, and analyse Web accessibility evaluation at a large scale. Complementary to these contributions, this thesis presents an implementation of accessibility evaluation that takes into account the problems tackled by this thesis in the subsequent Chapter.

The most beautiful thing we can experience is the mysterious. It is the source of all true art and all science.

Albert Einstein

Chapter 4

Characterising Web Accessibility at the Large

4.1 Introduction

Characterising the Web's accessibility is still a nascent research area, as discussed in the previous Section. It is known that developers and designers are largely unaware of accessibility practices, despite the efforts to educate and advise them with standards and guidelines such as WCAG and Section 508. Furthermore, when content production emerges as a common practice for end users, such as the case of Wikipedia, additional efforts should be taken, so that all content is accessible to people with disabilities.

In the light of these problems, this Chapter presents and discusses two experimental studies over accessibility quality of Web pages:

- *Wikipedia Web Pages*: this study centred on an initial foray on automated Web accessibility evaluation of a significant number of Web pages (over 7000), in order to understand the main problems of automating these evaluation practices, including potential scalability issues. Furthermore, this study envisioned the more practical study of understanding the impact of end-user content production in terms of accessibility;

- *Portuguese Web*: the second study centred on the analysis a large corpus of Web pages (over 20 million) from an accessibility perspective, which allowed us to reflect the findings of small-scale studies (such as the one conducted over Wikipedia), as well as shed some light on the characterisation of Web accessibility at the large.

The next Sections present these studies in more detail, as well as a discussion on the limits of Web accessibility evaluation automation practices.

4.2 Accessibility Evaluation of Wikipedia Web pages

Due to the decentralised nature of the Web's architecture, hyperlinks can be used as navigable and verifiable forms of making explicit the credibility of a Web page's content. Readers can follow these hyperlinks and judge by themselves whether the information they are reading is correct, especially since the linked information might have been created by authoritative and reputed sources. The establishment of this type of links is based on the assumption that reputed sources are trusted sources of confirmation from the point-of-view of readers.

Independent verifiability of human-created content is at the core of its credibility. Therefore it is of the uttermost importance to ensure that readers can access and understand information, otherwise it will be too cumbersome to verify it. While the content itself is at the core of credibility, the interface between content and the reader must be adequate. In other words, the readers must be able to *access* reputed sources without any kinds of barriers. For example, in a scenario where a Web page links to a supposedly authoritative source with a Web page that uses images to convey information, blind readers will not be able to grasp that information and, therefore, the credibility associated to this linking scenario is inherently flawed for this group of readers.

Wikipedia[1] is a free, editable-by-all online encyclopaedia. It has been a great phenomenon of massive collaboration of users, where exponential growth of the number of articles available (Voss, 2005) (around 2.5 million articles in English only, as of July 2008) is making it a highly used source of information.

The quality of the articles within Wikipedia is, therefore, the main aspect that drives users to browse and interact with it as a credible source for information (Hu *et al.*, 2007). Furthermore, Wikipedia's software platform, MediaWiki[2], delivers a collaborative environment which motivates contributors to iterate constantly on articles' contents, in order to improve their quality (Wilkinson & Huberman, 2007).

Next, this thesis details an experiment performed over the accessibility of Wikipedia articles and the external links embedded in them (i.e., references and citations). This experiment was formulated to verify the extent of hyperlinking and template mechanisms (sub-hypotheses *H1* and *H3*), as well as a smaller scale study of accessibility (sub-hypothesis *H4*).

[1] http://www.wikipedia.org
[2] http://www.mediawiki.org

4.2.1 Evaluation Methodology

To find the answers for these problems, an accessibility evaluation framework was created. This framework was based on the UWEM methodology (Velleman *et al.*, 2007), which provides directions towards the automation of the evaluation process (which becomes relevant due to scalability issues). The framework implements a subset of checkpoints from WCAG (the Web Content Accessibility Guidelines) that can assess particular accessibility issues of Web pages' HTML structures. Each checkpoint is responsible for analysing a specific detail on these structures (e.g., *image* tags without *alternate text* attributes). Afterwards, the outcome of each checkpoint contributes to a final, (quantitative) score representative of the accessibility level of the evaluated Web page. This experiment follows closely the setup presented below:

- $P = \{p_1, \ldots, p_{100}\}$, a set of 100 seed articles from Wikipedia, was randomly selected through a special purpose URL provided by Wikipedia[1];

- Each seed article's Web page, p_i, was analysed in order to extract all hyperlinks that point to Wikipedia articles, as well as other Web pages outside Wikipedia's domains. Thus, we excluded all pages from Wikipedia subdomains, such as hyperlinks to the article's equivalent in other languages, discussion pages, etc. A set $H_i = \{h_1, \ldots, h_n\}$ of Web pages was constructed based on this criterion;

- A pre-processing task was applied to each Web page (more specifically JTidy[2]), allowing the verification of standard-conformance (i.e., whether it respects HTML specifications), and pushing the application of healing processes on ill-formed Web pages (akin to what Web browsers do, as well), for structural analysis purposes;

- Finally, each Web page (both the initial set of pages and the external pages) was evaluated against a set of accessibility checkpoints $C = \{c_1, \ldots, c_n\}$. When combining the results yielded from this process into a quantifiable

[1] http://en.wikipedia.org/wiki/Special:Random
[2] http://jtidy.sourceforge.net

result A, we opted to apply a simple evaluation metric adapted from Sullivan & Matson (2000), presented in Equation 4.1:

$$A = \frac{1}{n} \sum_{i=1}^{n} c_i, c_i \in C \tag{4.1}$$

Here, each checkpoint c_i returns 1 if its corresponding assessment was successful when applied to the Web page that is being analysed, or 0 if not. Consequently, the overall evaluation result A averages all checkpoint values in the $[0, 1]$ range (i.e., a percentage), stating the accessibility quality level for the analysed Web page.

4.2.2 Results

By choosing 100 randomly selected root Web pages from Wikipedia, a total set of 7791 Web pages were crawled. Out of these, 100 were randomly selected, and 7691 hyperlinks were followed. From the followed hyperlinks, 7211 were targeted to other Web pages within Wikipedia, whereas 480 hyperlinks targeted to Web pages outside the scope of Wikipedia. Table 4.1 synthesises the overall results for accessibility assessment:

	Number	Total	%
Webpages correct	772	7725	9.99
Checkpoints correct	5	14	35.71

Table 4.1: Overall accessibility assessment

From these results, we verify that only nearly 10% of the Web pages fully complied with the checkpoints. We expect this situation to become worse when applying more complex and thorough evaluation procedures. Out of the 14 evaluation checkpoints analysed, only 5 were fully complied by all Web pages, based on UWEM tests: no *blink* elements, no *marquee* elements, no page refreshes through the *meta* element, the assurance that all *fieldset* elements have a *legend* child element, and the assurance that all *optgroup* elements have a *label* attribute.

When applying the accessibility metric defined above, we have an average quality of accessibility for each page of 84.6% ($\sigma = 0.088$). The minimum accessibility quality obtained from the evaluated pages was 50% (i.e., passing 7 checkpoints), whereas the maximum was fully compliance.

When splitting the analysis between internal and external Web pages, the results were different from the average values presented. Tables 4.2 and 4.3 further detail these findings:

	Number	Total	%
Webpages correct	750	7311	10.26
Checkpoints correct	5	14	35.71

Table 4.2: Internal Web pages accessibility assessment

	Number	Total	%
Webpages correct	22	414	5.31
Checkpoints correct	5	14	35.71

Table 4.3: External Web pages accessibility assessment

Analysing both tables, it becomes relevant to emphasise the fact that while the number of checkpoints passed remains the same, the number of Web pages which are fully compliant with all the 14 checkpoints evaluated is quite different between both clusters, with the ratio of 2:1 (*internal:external*). This is due to the fact that the overall HTML structure of Wikipedia complies with more checkpoints than the external Web pages linked from it. Moreover, the fact that Wikipedia provides a simplified markup language, might provide additional support to these values. Wikipedia's standardised environment facilitates the usage of accessibility aids that require well-structured Web pages. Based on these results, Table 4.4 presents a comparison between both Web page clusters, based on the accessibility metric defined above.

Once again, the ratio of compliance between Wikipedia Web pages and externally referenced Web pages differs almost by 8%, and the minimum expected quality for any Web page within Wikipedia is more than 7% higher than externally references Web pages.

%	Internal	External
Average	89.79	81.83
Maximum	100.00	100.00
Minimum	57.14	50.00
Standard deviation	8.68	9.66

Table 4.4: Internal vs. external Web pages accessibility assessment

Lastly, another analysis was performed based on clustering Web pages between the initially 100 seed Web pages and all the Web pages linked from these. Tables 4.5 and 4.6 present their respective summaries:

	Number	Total	%
Webpages correct	13	100	13.00
Checkpoints correct	8	14	42.86

Table 4.5: Seed Web pages accessibility assessment

	Number	Total	%
Webpages correct	759	7625	9.95
Checkpoints correct	5	14	35.71

Table 4.6: Child Web pages accessibility assessment

Both tables yield corollaries from the previous findings: if a random hyperlink is followed from a seed Web page, there is always the possibility of finding a Web page which cannot even be parsed (as child Web pages include externally linked Web pages). However, when comparing the accessibility evaluation between both clusters, as presented in Table 4.7, more results can be studied.

These numbers help on reflecting about the exploratory nature of interacting with large content, highly linked websites. If a user wants to follow a hyperlink from a Wikipedia Web page, there is no guarantee the linked Web page has a higher or lower accessibility quality. However, it is mostly expected that, due to the wilderness nature of the Web vs. the highly structured and templated nature of Wikipedia, that there is a concrete probability that the minimum quality of accessibility will decrease significantly.

%	Seed	Child
Average	84.14	84.65
Maximum	100.00	100.00
Minimum	64.29	50.00
Standard deviation	9.27	8.76

Table 4.7: Seed vs. child Web pages accessibility assessment

On Figures 4.1 and 4.2 these findings are further detailed. In both cases, the X-axis represents the accessibility quality level (in percentage, based on the accessibility metric previously defined) for Web pages, whereas the Y-axis represents the number of Web pages that meet those quality levels.

The shape of the graph for seed Web pages (i.e., Wikipedia articles) depicts the aforementioned stability of the accessibility quality. The "spike" centres around 80%, with a sharp decrease towards lower quality Web page count, and a more stable decrease on higher quality Web page count.

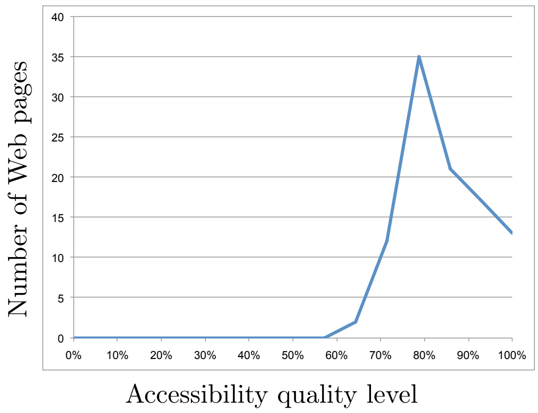


Figure 4.1: Accessibility quality distribution on seed Web pages

Regarding the shape of the graph for external Web pages (i.e., those used for referencing purposes within Wikipedia articles), a higher variation of quality is perceived. First, the initial spike shows that there is a high number of Web pages that do not meet the minimum requirements. This typically happens when pages are either totally inaccessible (e.g., HTTP 404 errors) or have severe parse errors. It is worth mentioning that the number of Web pages in this situation is 1.5 times more than the average Web page count for the large chunk of external Web pages analysed. From around 50%, the quality level distribution steadily increases until stabilising between 70% and 90%. Finally, the quality level decreases abruptly once again for the highest percentile (between 90% and 100%), which contrasts to the results of seed Wikipedia articles.

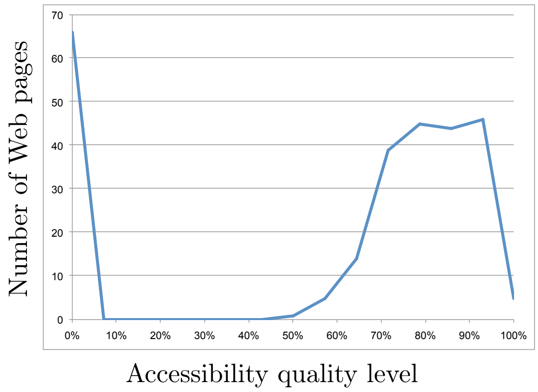


Figure 4.2: Accessibility quality distribution on external Web pages

Another aspect that has been analysed in this experiment concerns the number of errors and warnings yielded by the HTML parser (i.e., HTML wellformedness). No Wikipedia article processing has resulted on an error, which satisfies the high availability and good baseline quality of the MediaWiki software platform and its instantiation into Wikipedia. On the other hand, the high variability and uncontrolled nature of the Web has influenced the referencing mechanism used by Wikipedia article contributors. 66 external Web pages have resulted on errors, which accounts for nearly 25% of all external Web pages analysed. This aspect is manifested as the initial spike of the accessibility quality distribution presented in Figure 4.2.

The analysis of the seed Wikipedia articles has yielded 2323 warnings total, which accounts for an average of 23 warnings per article. In contrast, external Web pages have summed up to 19523 warnings total, on an average of almost 74 warnings per Web page. This 31.5% increase on warnings is another indicator of the accessibility discrepancies between Wikipedia articles and the Web pages used for citation and referencing purposes.

A deeper analysis of this part of the experiment is presented in the graphs depicted in Figures 4.3 and 4.4. Each of these graphs presents on the X-axis the number of warnings on a Web page (represented in a logarithmic scale), and, on the Y-axis, the number of Web pages that have those warnings.

Regarding the seed Wikipedia articles, the quick decay depicted on the graph in Figure 4.3 is directly correlated to the templating and stable mechanism of the MediaWiki software. No Wikipedia article that has been analysed had no warnings. Nevertheless, from 10 warnings onwards, the number of articles is close to zero.

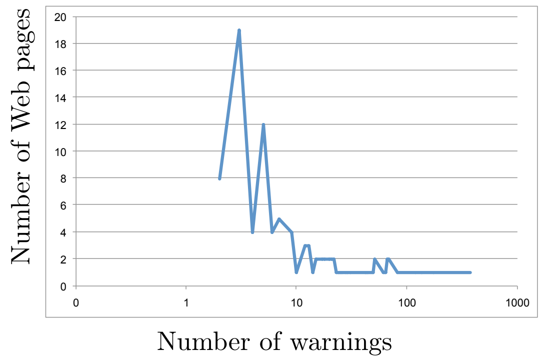


Figure 4.3: Warnings distribution on seed Web pages (log scale on number of warnings)

Regarding the distribution of warnings for external Web pages, the figure changes dramatically. The wilderness of the Web pays a significant role on the different distribution of warnings, which might be smoothed only by the influence of contributors on choosing the reference Web pages on Wikipedia articles.

The first interesting fact on this analysis is that there is a significant number of Web pages that have little to no warnings (in opposition to Wikipedia articles). However, this figure tends to shape differently afterwards. The biggest chunk of external Web pages orbit around 100 warnings, an order of magnitude above to those of Wikipedia articles.

4.2.3 Discussion

The main outcome of the conducted experiment concerns the high variability of the accessibility quality of external Web pages, i.e., ranging from no accessibility (or even unavailability) to an average distribution between 50% and 100%.

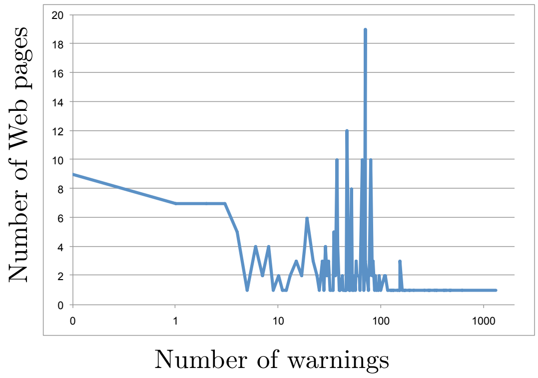


Figure 4.4: Warnings distribution on external Web pages (log scale on number of warnings)

Wikipedia provides a way to track unavailable links through the `{{dead link}}` template, resulting on more than 400,000 invalid links as of April 2007 (Wikipedia Contributors, 2011). However, in contrast to this situation, none is known about the accessibility of the available external links. Wikipedia does not provide similar templates to signal external Web page accessibility quality. This creates further accessibility problems, as Wikipedia contributors are unaware of the accessibility quality of the external Web pages they use for credible citation and referencing purposes.

Next, we detail another experiment on automatically evaluating the accessibility of Web pages, applied to a large set of Web pages from the Portuguese Web Archive.

4.3 Large Scale Web Accessibility Evaluation of the Portuguese Web

Based on the results and experience gathered in the previously experiment, this thesis discusses now a new experiment that was set up to assess a larger Web document collection provided by the Portuguese Web Archive. This experiment aimed at analysing the accessibility quality of the collection's Web pages, in order to understand if the raised hypothesis is verified. Next, a brief introduction to the Portuguese Web Archive is presented, followed by the conducted experiment and consequential results.

4.3.1 Portuguese Web Archive

A large amount of information is published exclusively on the Web everyday. However, after a few months, most of the published contents become unavailable and are lost forever (Gomes & Silva, 2006; Ntoulas *et al.*, 2004). The information published on the Web reflects our times and it is a valuable historical resource for future generations. For this reason, the Internet Archive (Kahle, 2002) has been collecting and archiving Web pages since 1996. Several countries have also founded archival initiatives for their own national Webs (Christensen, 2005; Philips, 2003).

The Portuguese Web Archive (PWA), a project of the Foundation for National Scientific Computing[1], periodically crawls Portuguese language contents, mainly from the *.pt* domain (Portugal), and stores them into a repository for archival purposes. Besides supporting research, the PWA has a direct interest in contributing to monitor and enhance the quality of the contents published on the Web. The accessibility quality of the archived contents is, therefore, an important measure which is being taken into account by the PWA.

4.3.2 Experiment

Since we predicted a certain level of complexity in this study, particularly due to the amount of data that would be analysed – a large collection of Web pages provided by the PWA –, we segmented the experiment into three main steps, as follows:

1. *Acquiring the Document Collection.* This first step centres on acquiring the collection of Web documents, by following specific crawling and storage processes;

2. *Accessibility Evaluation.* After obtaining the document collection, we will focus on computing data about the accessibility quality of each Web page, through a WCAG-based automated evaluation process;

[1]Foundation for National Scientific Computing: http://www.fccn.pt

3. *Data Analysis.* The third step focuses on computing accessibility metrics and other analysis processes on top of the evaluation results.

The following Sections further detail each step.

4.3.2.1 Acquiring the Document Collection

Collecting a representative sample of the Web for characterisation purposes is not a straightforward process (Henzinger *et al.*, 2000), due to both its size and its linking structure. However, national Webs, due to their smaller nature, are more treatable. Chung *et al.* (2009) show that national Webs tend to share the same macroscopic characteristics, independently from their internal structure differences. For these reasons, national Webs can be crawled and analysed in a practical and relevant way.

A crawler is a software that harvests content referenced by URLs and extracts its links for further processing, iteratively. While in theory this would allow for the collection of the entire Web, spider traps (Heydon & Najork, 1999) lead to corner cases on crawling (e.g., Web-based calendars generate an infinite number of Web pages). Thus, not every single Web page can be crawled and stored. For this reason, we established the following criteria as halting conditions within the crawling process (Miranda & Gomes, 2009):

- Maximum of 10,000 URLs crawled per Web site;
- Maximum content of 10 MB;
- Maximum number of 5 hops from the seed;
- Respect for the robots exclusion protocol (Koster, 1994);
- A courtesy pause of 2 seconds between requests to the same Web site (considering a *Web site* by its full qualified domain name).

To bootstrap the crawling process, we used as URL seeds nearly 200,000 site addresses hosted under the *.pt* Top Level Domain (TLD). The document collecting process was undertaken between March and May 2008. Most of the Web pages were collected during the first month, but the process was left running for two extra months reaching convergence. In order to create this Web document collection, the PWA developed a crawling system based on the following components, as detailed in Miranda & Gomes (2009):

- The Heritrix Web crawler (Mohr *et al.*, 2004) to collect Web pages;
- A software component to store and manage all crawled documents in the ARC file format, ideal for the archival of large collections of Web documents (Burner & Kahle, 1996).

Due to the way these document collections were created and maintained (especially due to its large size), the architectural aspects that allow evaluation processes to scale freely limits these algorithms to HTML inspection. One of such instances relates to associated resources – images, external scripts, applets – and its inclusion on the evaluation process.

4.3.2.2 Accessibility Evaluation

After the creation of the collection of Web pages, we developed software components to process it according to automatable Web accessibility evaluation practices. Since the evaluation of large Web document collections can be a resource intensive process (i.e., CPU, memory, disk), we implemented the evaluation software in order to minimise its impact resource-wise. We opted for an implementation of a set of evaluation techniques from WCAG 1.0 (Chisholm *et al.*, 1999). The rationale for version 1.0 of these guidelines instead of the more recent version 2.0 is due to two main reasons: first, the implementation of WCAG 1.0 checkpoints can be easily supported by existing Web accessibility evaluation practices, in order to ensure their correctness; and second, the crawling process and subsequent document collection creation were performed before WCAG 2.0 achieved a *W3C Recommendation* status.

We used the hardware and software infrastructure created by the PWA for processing large document collections. On the hardware side, 10 blade servers were made available for this experiment, each with 2x Quad-core 2.3GHz CPUs and 8GB of RAM. Regarding software components, we developed an evaluator for WCAG 1.0 that sits on top of Hadoop (Apache Foundation, 2010), an open-source implementation of Map/Reduce (Dean & Ghemawat, 2004), leveraging the hardware cluster's scalability properties.

We implemented the techniques for 39 checkpoints from WCAG 1.0, focusing on HTML structure analysis of the collected Web documents, using UWEM (Velleman *et al.*, 2007) as the benchmark implementation reference. Consequently, all checkpoints are either from priority 1 or 2, thus leaving out the most advanced – albeit less critical – checkpoints from priority 3. Following UWEM, we categorised the expected results of evaluating a checkpoint as follows:

- *PASS*: the checkpoint is applicable to an HTML element *and* its compliance is verified automatically;
- *FAIL*: the checkpoint is applicable to an HTML element *and* its compliance unachieved;
- *WARN*: the checkpoint is applicable to an HTML element *but* it is impossible to verify its compliance.

UWEM suggests an implementation of checkpoints with a mix of XPath 1.0 (Clark & DeRose, 1999) expressions and verbose text. However, using XML-based technologies such as XPath put difficulties on parsing malformed HTML documents, which are significant in any Web document collection, as well as performance penalties when scaling up the evaluation process. Consequently, we have opted to implement our accessibility evaluation component in the Groovy programming language[1], aided by the NekoHTML parser[2]. Groovy's native capabilities of writing terse source code allowed us to migrate the XPath expressions defined in UWEM into equivalent native code expressions, thus maintaining the semantics defined by UWEM evaluation, as well as build upon the optimisation techniques of the Java Virtual Machine.

The main evaluator component implements all checkpoints and applies them to all HTML elements of each Web page in the collection. In order to minimise the impact on CPU and memory and, consequently, comply with the desired

[1]Groovy Programming Language: http://groovy.codehaus.org/
[2]NekoHTML Parser: http://nekohtml.sourceforge.net/

scalability for this study, we took advantage of streaming and caching algorithms in the implementation of the Web accessibility evaluator. All evaluation results were collected and stored for data analysis tasks, as detailed next.

4.3.2.3 Data Analysis

Having the raw results of the accessibility evaluation, we processed data in order to have answers for our research question. For this task, we used the R Statistical Computing framework[1] and ancillary shell scripts. To assess a quantification value of the accessibility quality of Web page, we defined three evaluation rates based on the *failure rate* metric by Sullivan & Matson (2000).

We opted to adjust this metric by taking into account the amount of times a checkpoint fails in the same Web page. This option was based on the assumption that a Web page that fails once must yield a smaller penalty than if it fails several times. Each rate is normalised into a percentage, with the semantics from *not accessible* (0%) to *fully accessible* (100%). Based on these conditions, we defined our automated evaluation process and corresponding metric rates with the distinction of whether a checkpoint evaluation of a given HTML element results in *PASS*, *FAIL*, or *WARN*, as follows:

- **Conservative rate**: *WARN* results are interpreted as *failures*. The semantics of this rate conveys the worst-case scenario on accessibility evaluation:

$$rate_{conservative} = \frac{passed}{applicable} \tag{4.2}$$

- **Optimistic rate**: *WARN* results are interpreted as *passed*. This rate is related to a best-case scenario where developers and experts dismiss warnings – often incorrectly, as explained in Sloan *et al.* (2006) –, as accessibility issues that were taken into account:

$$rate_{optimistic} = \frac{passed + warned}{applicable} \tag{4.3}$$

[1] R Statistical Computing: http://www.r-project.org/

- **Strict rate**: *WARN* results are *dismissed*, thus accounting only the actual *FAIL* results:

$$rate_{strict} = \frac{passed}{applicable - warned} \tag{4.4}$$

Each rate was computed for each checkpoint, as well as their aggregation into a final score stating the accessibility quality of a Web page. Next, we present the results of this experiment.

4.3.3 Results

The evaluation of the Web document collection yielded results for 28,135,102 HTML documents, out of 48,718,404 Web content documents (Web pages, images, PDFs, etc.) that have been crawled in total (nearly 58%).

A total 40,831,728,499 HTML elements were analysed, implying an average 1451 HTML elements per Web page. Of these, 1,589,702,401 HTML elements successfully met all applicable Web accessibility criteria, an average of 56 HTML elements per Web page (around 3.89%). Regarding failures, 2,918,802,078 HTML elements failed to comply, corresponding to an average of more than 103 errors per Web page (approximately 7.15%). Finally, 36,323,224,020 HTML elements were detected for belonging to a *warning* status, accounting for an average 1291 warnings per Web page (nearly 89%).

These numbers provide a rough characterisation of the document collection, the evaluation process, and of the overall quality of accessibility on the Web. But to get the gist of why is accessibility quality so low on the Web, we analysed all the computed data by focusing on distributions of the evaluation data. The next Sections present a more detailed set of results in the following fronts: *distribution of rates*, *rates and page complexity*, *rates and hyperlinking*, and *alternative texts*.

4.3.3.1 Distribution of Rates

We aggregated the accessibility quality rate by permillage values, and generated a frequency distribution of how many Web pages belong to each aggregation. Figure 4.5 presents the distribution of *conservative rate* metric versus page count. Since all *warnings* are interpreted as errors, and no Web page was missing the

HTML elements detectable in the checkpoints that yield warnings, no Web page was able to reach the maximum value of accessibility quality. The depicted exponential decay starts around 5% of compliance, where the number of pages with good quality is minimal.

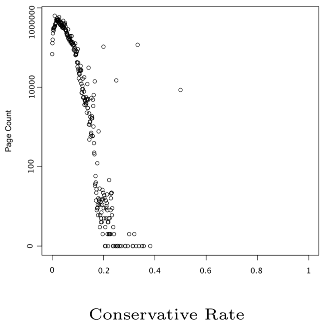


Figure 4.5: Accessibility distribution for *Conservative rate*

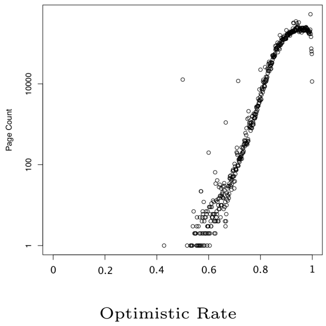


Figure 4.6: Accessibility distribution for *Optimistic rate*

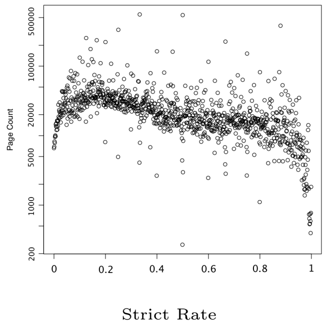


Figure 4.7: Accessibility distribution for *Strict rate*

Figure 4.6 presents the distribution of *optimistic rate* metric versus page count. Since this metric takes into account all *warnings* as positively complied, all checkpoints that cannot be univocally evaluated have a significant positive effect on its page count distribution. Here, we observed that there is a rapid progression of the number of pages for each aggregated rate, with a lower bound of accessibility quality around 50% and a 90% mean value.

When analysing from the perspective of the 100% detectable problems (i.e., just *errors*), we found that the there is a near constant distribution of Web pages according to their accessibility quality, as depicted in Figure 4.7. The only exceptions on this are at the edges of the distribution, especially when approaching fully compliance with the detectable errors. Here, the decay on the page count is still measurable, despite the fact that it is less steep comparing to *conservative rate*.

4.3.3.2 Rates and Page Complexity

Our second incursion on this study relates to the verification of a correlation between the rate and complexity of each Web page, i.e., the number of HTML elements present in a Web page, encompassing both the breadth and depth of the Web pages' HTML element tree, as presented in the next Figures (log-scale on page complexity – Y-axis).

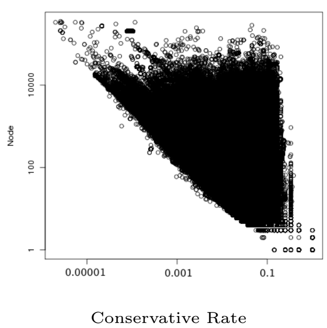


Figure 4.8: Accessibility *conservative rate* versus page complexity

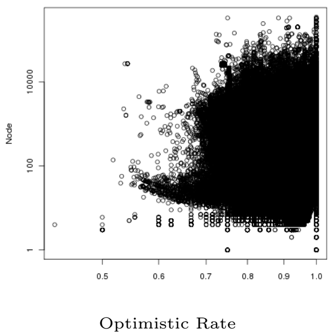


Figure 4.9: Accessibility *optimistic rate* versus page complexity

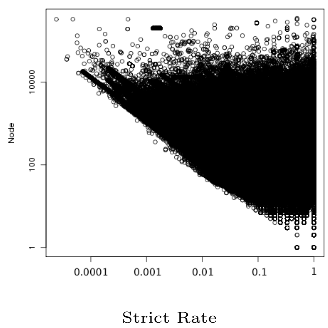


Figure 4.10: Accessibility *strict rate* versus page complexity

Regarding *conservative rate*, with the exponential decay of HTML elements count, the accessibility rate approaches 10% quality, as presented on Figure 4.8. When taking into account the *optimistic rate* metric, there is no obvious correlation between elements count and accessibility quality, as depicted on Figure 4.9. Nevertheless, there is a homogeneity on the distribution of *optimistic rate* regarding elements count. Lastly, Figure 4.10 depicts the distribution for the *strict rate* metric. Like in the *conservative rate* metric, we discovered the same kind of exponential decay between elements count and the metric. However, in this case, the rate approaches 100%, since *warnings* were not taken into account. A comparison between all rates can be visualised next in Figure 4.11.

4.3.3.3 Rates and Hyperlinking

As discussed earlier, the typical accessibility evaluation of Web pages is performed over a single Web page or a Website. While this type of measurement is important by itself, it does not take into account one of the core aspects of the Web – *hyperlinks*. Based on this assumption, we analysed the accessibility quality of Web pages from the perspective of their *outdegree* – how many outbound links does a Web page contain, as presented below.

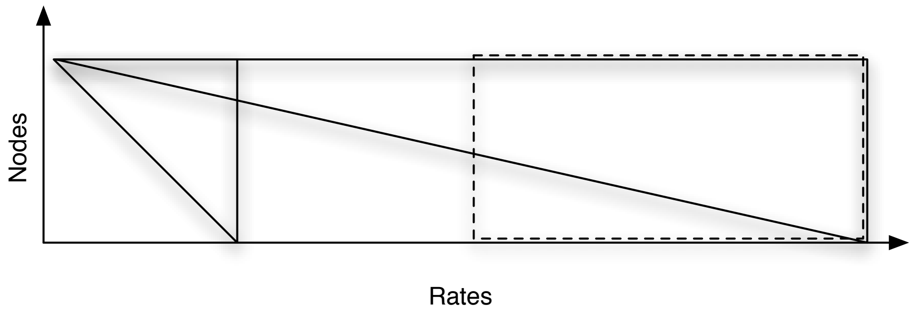
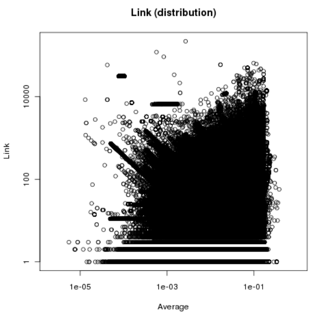
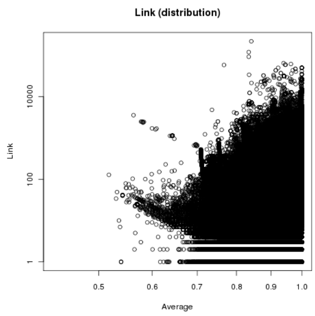


Figure 4.11: Comparison between *Rates*: *conservative* (left triangle), *optimistic* (dotted rectangle), and *strict* (right triangle)



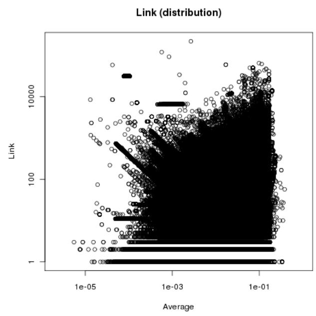
Conservative Rate

Figure 4.12: Accessibility *conservative rate* versus outdegree



Optimistic Rate

Figure 4.13: Accessibility *optimistic rate* versus outdegree



Strict Rate

Figure 4.14: Accessibility *strict rate* versus outdegree

This resulted on finding out that Web pages having few links tend to have a smaller accessibility quality, a property that manifests similarly in the three rates. When accessibility quality improves, the variability of the outdegree increases proportionally. In other words, Web pages with a high number of links have a high variability on their accessibility quality, but those with few links tend to have their accessibility quality worsened.

4.3.3.4 Alternative Texts

While it is important to characterise Web accessibility at the large based on consolidating the results from all checkpoints, it is also relevant to analyse the shape of single checkpoints at a large scale. In this Section we discuss one particularly important checkpoint from WCAG 1.0, Checkpoint 1.1, which states: *"Provide a text equivalent for every non-text element (e.g., via* alt*,* longdesc*, or in element content)"*.

By principle, all checkpoints are of uttermost importance, for accessibility guideline compliance. However, Checkpoint 1.1 deals with both *the ability to access information* – accessibility on its stricter sense (e.g., people with blindness are unable perceive the information depicted on an image) – and affords an easily detectable success criteria (*alt* or *longdesc* attributes, or in element text).

We applied the same three rate metrics to the results from WCAG 1.0 Checkpoint 1.1. Here, we also wanted to understand the distribution of the rates, and how it maps into Web page count for each rate aggregation, as presented in the next Figures (log-log distribution).

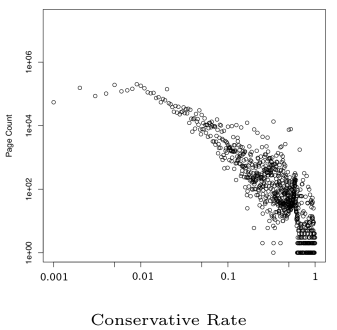


Figure 4.15: Compliance with WCAG Checkpoint 1.1 based on *conservative rate* evaluation

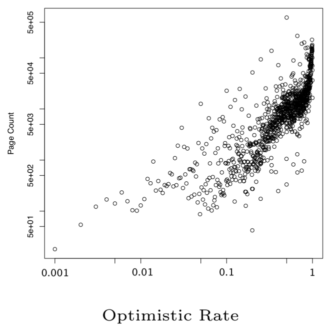


Figure 4.16: Compliance with WCAG Checkpoint 1.1 based on *optimistic rate* evaluation

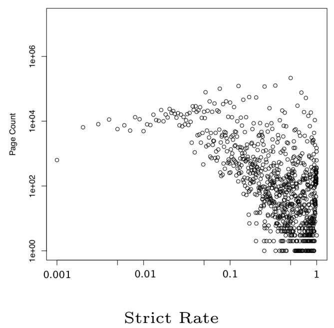


Figure 4.17: Compliance with WCAG Checkpoint 1.1 based on *strict rate* evaluation

In the case of the *conservative rate* distribution, as depicted in Figure 4.15, we verified that page count decreases exponentially with the increase on Checkpoint 1.1 compliance, closely resembling a power-law distribution. When going into the *optimistic rate* distribution, the opposite situation occurs. There is an exponential increase between page count and the rate distribution due to *warnings* being accounted as positive assertions of Checkpoint 1.1, as depicted in Figure 4.16. Lastly, regarding the *strict rate* distribution, *errors* are distributed in a less steep path in comparison with the *conservative rate*, as presented in Figure 4.17. The page count variability is stable until quality rate reaches approximately 5%, where page count variability increases proportionally to the quality rate.

4.4 Discussion

One of the interesting aspects arising from this large-scale experiment is the distribution of *conservative*, *optimistic*, and *strict* rates. When looking at *errors* distribution (i.e., *strict rate* metric), its linearity implies that critical accessibility problems are likely to be encountered with the same probability by end users who depend on proper accessibility. However, when taking into account *warnings*, the picture of accessibility on the Web is not clear. When *warnings* are perceived as positive, accessibility quality quickly reaches high levels. But, as discussed by Vigo *et al.* (2007), it has been observed in small scale that the rate of warnings goes hand in hand with error rate. This result has the direct consequence that indeed the results from such small scale studies are verified at the large scale as well.

Another important result from this experiment concerns the relationship between the number of HTML elements in a Web page and its accessibility quality. While applying the *optimistic rate* is insufficient to reach a significant conclusion, this changes in what respects to both *conservative rate* and *strict rate* . In both cases, we verified that Web pages with a high number of HTML elements are correlated with their accessibility quality. There was no single Web page in the evaluated document collection that had both a small HTML element count and a poor accessibility quality rate. We hypothesise that this happens due to the complexity of Web pages: simplicity leaves out several HTML structural compositions that hinder accessibility, and also that a smaller Web pages are more manageable while designing/implementing them.

Finally, when going deeper into this evaluation, Checkpoint 1.1 also provides some clues about the state of accessibility on the Web. This checkpoint is constantly being used as an example poster child for accessibility issues, since it is easy to conceptualise the problem behind the checkpoint. Even so, the distribution of its compliance rates implies that the message is not being disseminated enough. In this checkpoint, *warnings* signaled the impossibility of automatically inferring if there is a textual equivalent for a given non-textual element.

But for those cases where it is possible to algorithmically decide that a media element does not have a textual equivalent (e.g., the *alt* attribute on an *img* element), the distribution for the *strict rate* also denotes a strong decay in this checkpoint's compliance. Based on this distribution, we argue that Web page creators still have a difficulty on providing accessible contents, even if there are direct mechanism within the HTML specification to encompass this aspect. Consequently, based on the state-of-the-art of accessibility compliance, we believe that the distribution for *conservative rate* on Checkpoint 1.1 is close to reality. Recent advancements in textual analysis of alternative texts, as described by Olsen *et al.* (2010), with the aid of machine-learning techniques, are starting to opening the way to better understand this type of issues on evaluating Web accessibility compliance.

4.4.1 Impact on Designing Accessible Web Pages

The results of our experiment can also be discussed towards more practical matters, i.e., how people who create Web pages (e.g., designers, developers, etc.) can mitigate the recurring accessibility problems encountered on the Web.

Paying attention to detail in the structure, rhetoric and discourse of a Web page conveying information is critical for its accessibility success. The *warnings* raised by the evaluation process concern the lack of usage of HTML structural elements that help building the discourse of a Web page. Therefore, we believe that there is a strong need for a better education and dissemination of best practices for properly using the semantics of HTML elements.

Another issue concerns the aforementioned problem of the relationship between Web accessibility quality and the complexity of Web pages. Our position on this issue, in what respects to designers and developers, is that Web accessibility is more manageable in *smaller chunks*. Our advice for Web page creators is to follow a *simplicity* approach on defining the structure of Web pages, which lowers the burden of verifying accessibility compliance during development. This goes hand in hand with the usage of templates for lowering the burden of maintaining the quality of a Web page.

4.4.2 Impact on the Perception of Accessibility

Our results show the profound difference between the opposite perspectives of accessibility given by the *conservative* and *optimistic* rates. Overall, the *conservative* results are in pair with the *strict* analyses performed.

On the other hand, when comparing with the *optimistic* rate, it shows how developers and designers might interpret the accessibility quality of the Web sites they create, i.e., having an *optimistic* view. Since most developers and designers are not accessibility experts, and since non-experts tend to incorrectly evaluate accessibility (Yesilada *et al.*, 2009), we hypothesise that the *optimistic* rate might shed light on the real perception of accessibility by non-experts at the large. As presented in the Results Section, around 90% of the HTML elements that were evaluated yielded a *warning* result. Consequently, most of the results from evaluation will be erroneously interpreted by non-experts.

This discrepancy further shows that guidelines are just starting points for proper accessibility adequacy, i.e., they must be complemented with in-depth user testing. Furthermore, we believe that improvements must be made on communicating guidelines and presenting accessibility evaluation results to motivate developers and designers on investigating the nature of accessibility evaluation *warnings*.

4.4.3 Hyperlinking and Accessibility

We believe that accessibility on the Web must be analysed not just on a single Web page (or Web site) perspective, but also by taking into account the relationship between Web pages. This distinction separates existing Web accessibility evaluation studies (microscopic scale) from a Web Science perspective on Web accessibility (macroscopic). Earlier in this Chapter we explored one important factor of hyperlinking, the *outdegree* of each Web page and analysed the relationship with its accessibility quality.

Dill *et al.* (2002) discovered that the Web follows a fractal structure in the way Web pages are linked between each other: strongly connected clusters of Web pages (such as national Web graphs) function as both inbound and outbound hyperlink hubs. The frequency distribution of these connections is scale-free – most Web pages have few links, and are linked by few Web pages – a feature greatly exploited by Web page ranking algorithms (Brin & Page, 1998).

The *outdegree* of a Web page reflects if its content tends to be generated (high outdegree) or hand-made (low oudegree), as pointed out by Baeza-Yates *et al.* (2007). Therefore, this metric can be used as a compass to understand whether if hand-made Web pages tend to have a higher accessibility quality, in comparison to generated content (e.g., with the aid of content management systems).

Based on these assumptions, our results show that in fact, the outdegree of a Web page provides some guidance on characterising accessibility at the large. For Web pages that have a high number of hyperlinks, the accessibility quality is higher. It is worth noticing that this correlation must not be interpreted as cause-effect. In contrary, it does show that using mechanisms such as HTML template languages (e.g., used in content management systems) help improving the accessibility of Web sites. Our experiment reveals this pattern at the large scale, which we already shown at a smaller scale studies of the accessibility of Wikipedia (Lopes & Carriço, 2008).

The other side of this analysis also provides some insights on the applicability of Web accessibility guidelines by designers and developers. For Web pages that had a low outdegree, we found that they had a high variability on their accessibility quality. This fact reflects the disparity on the expertise of designers and developers: those who hand code Web pages (i.e., detected through the tendency towards lower outdegrees) are part of a diverse group of people with different knowledge about accessibility.

4.5 Summary

This Chapter presented two experimental accessibility evaluation studies of the Web at different scales. A first, small-scale study centred on Wikipedia investigated the state of accessibility of Wikipedia articles versus externally linked Web pages (used for citation and information credibility purposes). The second, large scale study of accessibility on the Web conducted over a Web document collection provided by the Portuguese Web Archive.

The results of these study revealed a set of characterisations of Web accessibility, both on how it impacts users and how Web page creation hinders it. We have discovered the effects of Web page quality in what respects to accessibility, and how it hinders the expected universality aspects of the Web. One of the aspects studied leveraged the confirmation that simpler, smaller Web pages tend to have a better accessibility quality. We hypothesise that is due to providing less margin of error for Web designers and developers. Our results also show that accessibility communication must be further improved. This was also shown through the poor compliance levels of WCAG Checkpoint 1.1 (alternative texts for media elements), as well as on the disparity between *conservative* and *optimistic* perspectives over Web accessibility evaluation results. This study also uncovered how the fundamental hyperlinking properties of the Web differ when taking account accessibility.

We have no idea about the real *nature of things... The function of modeling is to arrive at descriptions which are useful.*

Richard Bandler and John Grinder

Chapter 5

Modelling Distributed Web Accessibility Evaluation

5.1 Introduction

Web Science has emerged as the discipline that tackles the Web as a core object of study. This vision poses new challenges on how to perceive the impact of technology and society on the Web and vice-versa. At the same time, the Web Accessibility Initiative defines a particular set of methods and technologies to improve the experience of users with disabilities while interacting with the Web. As discussed in Chapter 2, there is still a gap on research between both fields. In this Chapter we propose a new model for research on Web accessibility centred on evaluation practices, in order to tackle macroscopic properties of Web accessibility. The basis of this model lays at the findings, experience, and insights gathered from the experiments presented in the previous Chapter.

5.2 A Framework for Web Accessibility Research

5.2.1 Fitting the WAI Model to the Web Science Lifecycle

At the core of the Web science framework lays its lifecycle, comprising on the dichotomy between the microscopic and macroscopic scales of analysing and understanding the Web. Therefore, to properly frame the Web accessibility discipline into the methods of Web science, it is first necessary to understand how the WAI model fits itself into this lifecycle. The WAI model provides a set of engineering solutions to cope with Web accessibility issues, thus mapping themselves into microscopic scale of the Web science lifecycle, as depicted on Figure 5.1.

The lifecycle starts with the set of *issues* tackled by Web accessibility, i.e., the fact that many Web sites contain several barriers for people with disabilities. To mitigate this problem, WAI defines a process based on the creation of Web accessibility guidelines. This process drives *social* solutions in the form of aiding developers producing accessible content (i.e., through WCAG), as well as a more *technical* side of tools to help creating and interacting with accessible content (i.e., ATAG and UAAG). Both sides, *social* and *technical* lead the way to the creation and interaction of accessible Web sites.

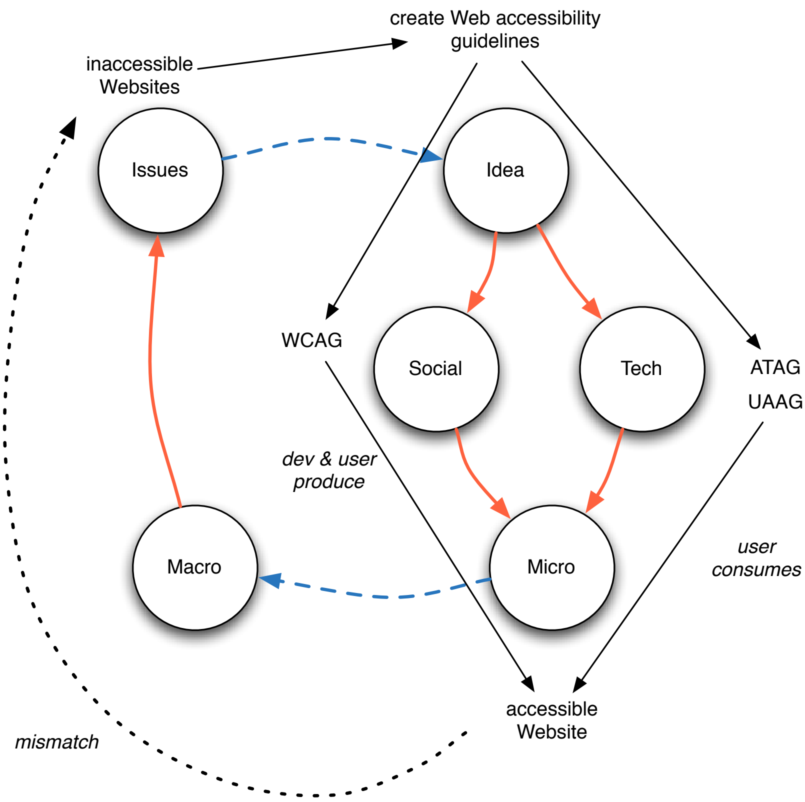


Figure 5.1: Decomposition of the WAI Model into the Web Science Lifecycle

However, despite the fact that these guidelines and processes exist, the focus given on accessibility compliance and evaluation from the WAI is still centred on microscopic levels, e.g., developers and designers and the corresponding Web sites they are focused at. This lack of analysis on the macroscopic effect of such processes results on a *mismatch* with the goals of Web science.

The macroscopic effect of non-compliance with the WAI process results on an inaccessible Web, as demonstrated in the previous Chapter. It is known that some factors influence this tendency, such as developers not having knowledge about Web accessibility, product managers not planning resource allocation for Web accessibility compliance, etc. However, little is known about what effects the WAI model has at the scale of the Web, e.g., if it reflects an accessible subset of the Web, following the same connectivity properties of the Web. Having this knowledge will open the way for the Web accessibility research and education community to tailor technical and social practices towards improving the Web for people with disabilities.

In order to understand what are macroscopic properties of Web accessibility, it is first necessary to grasp the (non) compliance of Web sites towards WCAG. Since a single Web site (or Web page) does not entail all the diversity of the Web, there is a need to first choose a representative collection of Web sites and, afterwards, evaluate their WCAG compliance, analyse these results, and extrapolate macroscopic properties. For an effective collection and analysis process, we argue that WCAG-based evaluation processes must scale to accommodate the Web's dimensions, as detailed next.

5.2.2 Scaling Web Accessibility Evaluation

Evaluating the accessibility of large collections of Web pages representative of the diversity of content on the Web is a complex task. First and foremost, expert evaluation by hand can hardly scale beyond tens of Web pages, as it is a time consuming activity. Therefore, the evaluation of Web accessibility becomes more effective with its automation.

Automated Web accessibility evaluation is a well-known process of getting the gist of a Web page's accessibility quality, typically based on the outcome of

WCAG compliance check (Abascal *et al.*, 2004). This testing done based on miscellaneous techniques defined by WCAG, such as HTML inspection, colour checking, etc. However, due to the complexity of such analysis techniques (e.g., image processing (Bigham, 2007)), existing automated evaluation software is typically a partial implementation of the defined compliance checking procedures (Aizpurua *et al.*, 2009b).

On the other hand, while this process is orders of magnitude faster than manual evaluation, the sheer size of the Web puts a high burden on grasping the accessibility quality of large Web page collections. To mitigate this issue, several methods for selecting random samples of Web pages have been created in order to increase automated evaluation throughput (Brajnik *et al.*, 2007; Velleman *et al.*, 2007). However, these techniques have been defined to extract a subset of Web pages of single Web sites. It is known that the Web's diversity is mapped in a fractal way (Dill *et al.*, 2002), which requires more expedite and thorough sampling methods, such as collecting national Webs (Baeza-Yates *et al.*, 2007).

Consequently, a *divide-and-conquer* approach to the evaluation of Web accessibility at a large scale will facilitate specially in terms of computational resources required. We delineated this approach in two orthogonal domains: *technology* and *topology*. Next, we detail each one of them.

5.2.2.1 Technological Evaluators

The analysis of Web accessibility evaluators presented earlier in the State-of-the-Art Chapter shows that there is a trend of Web accessibility evaluators centred on the following *default profile* dimensions:

> *Automated* evaluation of *single pages*, implementing *HTML* analysis based on *WCAG 1.0*, generating human-readable *reporting*, made available through a *Web interface* with a *freeware* license.

It is clear that this profile provides insufficient achievements towards the goal of a distributed model for technological evaluators of Web accessibility quality. Since each of these dimensions has a different stake and implication in the definition of the distributed model, we separated them into the following categorisation:

- *Core dimensions*: to support the distributed model, evaluators must comply with a specific set of assumptions – they should perform *automated evaluation* of *Web pages* and generate *machine-readable reports*, and allow for their usage as a *Web service* or afford the replication of this process (e.g., *open source server*);

- *Domain-specific dimensions*: evaluators should target at least one *Web page technology* (e.g., HTML), based on one or more *Web accessibility guidelines.*

Based on this categorisation, we defined a model for distributed evaluation of the accessibility of Web pages, as depicted next in Figure 5.2:

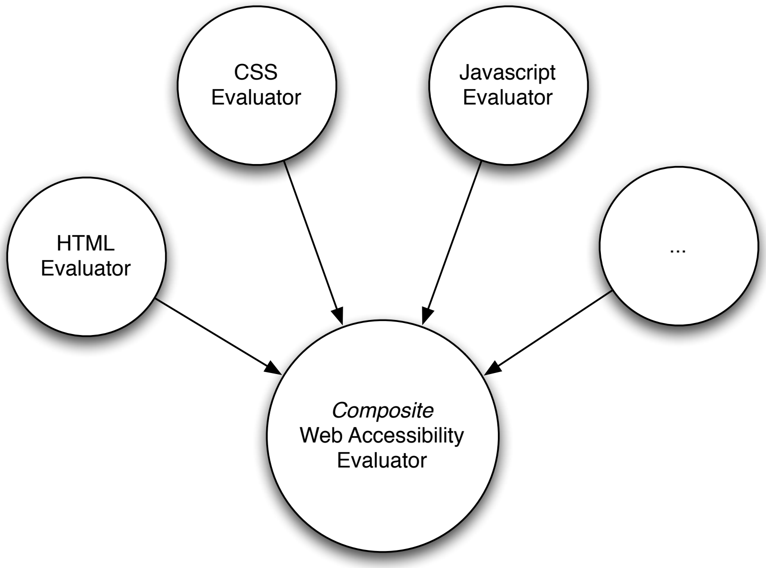


Figure 5.2: Distributed Technological Evaluators for Web Accessibility

This model is based on having *atomic* evaluators targeted at a client-side technology and an accessibility guideline, such as *HTML techniques for WCAG 2.0 evaluation.* These evaluators are, then, combined into more complex *composite* technological evaluators which, ultimately, provide a thorough verification of the accessibility quality of a Web page. To accomplish this model, each evaluator must comply with the *core dimensions* defined above.

It is known that some evaluation techniques are harder to implement in an automated fashion (Vigo *et al.*, 2007), which leads to limitations in the way evaluators comply with conformance checking of accessibility guidelines. Our model for distributed evaluation takes this issue into account in two fronts, *Shared Atomic Evaluators* and *Multi-layer Composite Evaluators*, as depicted next in Figures 5.3 and 5.4:

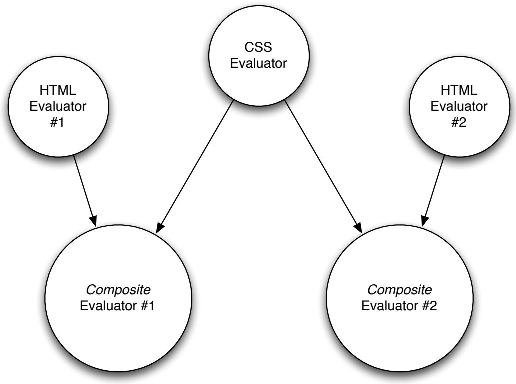


Figure 5.3: Shared Evaluators

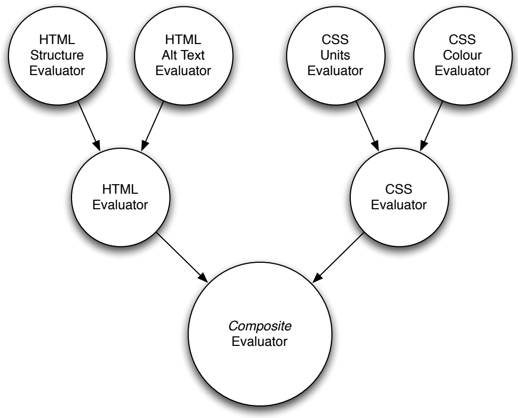


Figure 5.4: Multi-layer Evaluators

The availability of a technological evaluator opens the door not just to their integration into a composite evaluator, but also more flexible integrations. Novel technology evaluation techniques appear all the time, such as complex analysis of CSS (Aizpurua *et al.*, 2009a) or alternative text analysis with machine-learning techniques (Bigham, 2007; Olsen *et al.*, 2010). By having programmatic access as a *Web service* and a *machine-readable reporting* format, composite evaluators can benefit from this kind of techniques seamlessly. In the case of *shared evaluators* (Figure 5.3), atomic evaluators are integrated into more than one composite evaluator, such as affording CSS analysis techniques for evaluators targeted at different accessibility guidelines. Whereas, for *multi-layer evaluators* (Figure 5.4), technology evaluators themselves can be defined as compositions of several evaluators which can be either composed or atomic. In sum, with this model for distributed technological evaluation of Web accessibility, we aim to:

- Lower the burden of implementing Web accessibility evaluators;

- Connect accessibility evaluation for the technologies present in Web pages;
- Increase the availability of *novel evaluation techniques* for accessibility guidelines;
- Foster the scalability and distribution of automated Web accessibility processes.

5.2.2.2 Topologic Evaluators

Due to the sheer size of the Web, in number of sites and pages, it is impractical to study its accessibility properties in large scale without having a huge number of *CPU-hour* resources (Lopes *et al.*, 2010a). Ideally, this problem could be mitigated by having each Web server describe the accessibility quality of the Web sites it contains. It would require the installation and configuration of a *technological evaluator* and make *reporting* available upfront. However, this solution is unrealistic due to miscellanea factors such as the lack of knowledge on accessibility by several stakeholders.

A large-scale analysis of the Web's accessibility properties must take into account these limitations. Therefore, we devised a model for distributed evaluation of Web accessibility which relies on the Web topology for *domain names*, akin to how the domain names resolution protocol is performed (Mockapetris, 1987). This model defines the following set of rules to accomplish a distributed Web accessibility evaluation:

- Domains are represented by a *topological* evaluator, which answer queries about the accessibility for a given URL;
- Each top-level domain (TLD) has available a set of *technological* evaluators to handle any domain under its supervision;
- All domains must know which descendant domains have a set of *technological* evaluators;
- When queried for an accessibility evaluation for a specific URL, *topological* evaluators must navigate through the domain tree until reaching the appropriate *topological* evaluator and associated *technological* evaluators.

While simple in its form, this model provides support for several use cases on Web accessibility evaluation, such as:

- TLDs can support accessibility evaluation for all the domains under their control;
- Any domain can have its own *technological* evaluators;
- Specific domains can be monitored for their accessibility compliance (e.g., governmental bodies);
- Different domains can be supported by different *technological evaluators* targeted at specific accessibility guidelines;
- Scalability of Web accessibility evaluation can be reached by making available as many *technological* evaluators as necessary, bounded to different domains in the *topological* evaluators domain tree.

An example for the distributed topological evaluators model is presented next in Figure 5.5:

This example depicts some of the aforementioned use cases, in the context of the *.pt* TLD. This TLD has available two default *technological* evaluators to assess the Web accessibility quality under its management. Since the *.com.pt* and *.edu.pt* do not have their own evaluators, the default *.pt* evaluators are used to assess all descendent domains. However, the *.gov.pt* domain has two evaluators available, which can be used to assess all governmental Web sites. Lastly, the *infosociety.gov.pt* Web site uses its own *technological* evaluator to monitor the accessibility of its Web pages.

5.2.2.3 Mixing Evaluators

Both *technological* and *topological* models provide helpful support for a distributed Web accessibility evaluation. However, some issues arise with regards to the perfusion of *technological* evaluators when taking into account the distributed *topological* model:

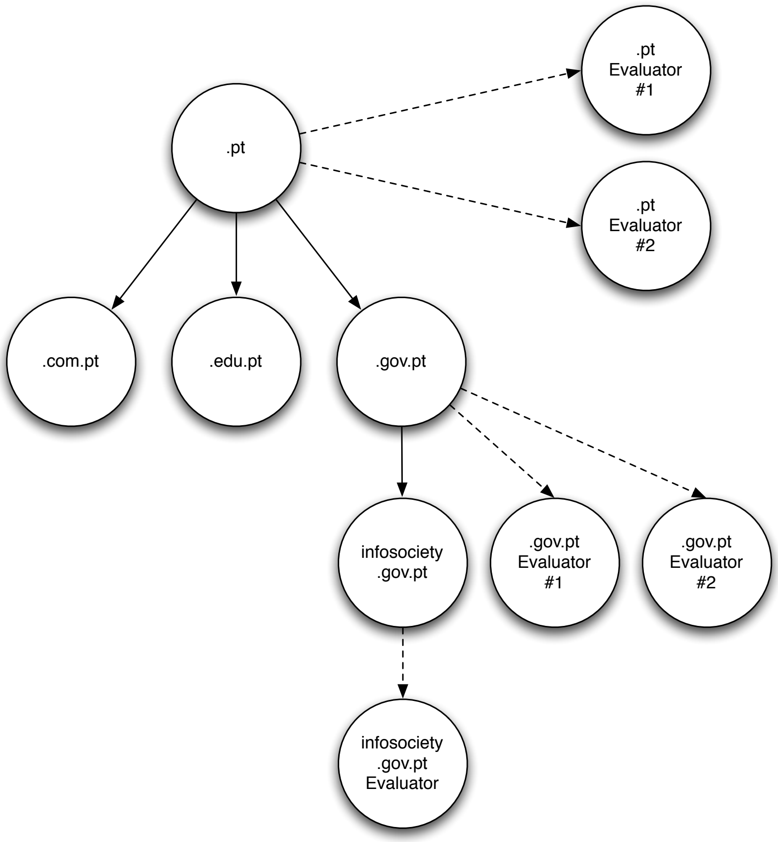


Figure 5.5: Distributed Topological Evaluators for Web Accessibility

- New evaluation technologies are constantly appearing, which pose a problem in their perfusion across *topological* evaluators;

- *Atomic* evaluators can be computationally intensive, requiring substantive available resources for their execution (e.g., verifying if images are colour-blindness friendly).

For this reason, we explored the capabilities provided by both distributed evaluation models. As detailed earlier in the *technological* model, *atomic* evaluators can be shared between several *composite evaluators*. Since the *topological* model depends on the existence of *technological* models, a by-product of mixing both models is the possibility of integrating *atomic* evaluators in different *topological* models. In Figure 5.6 we present an example combination of both models, where a single *alternative text* evaluator is leveraged by different evaluators of the *.pt* and *.gov.pt* domains:

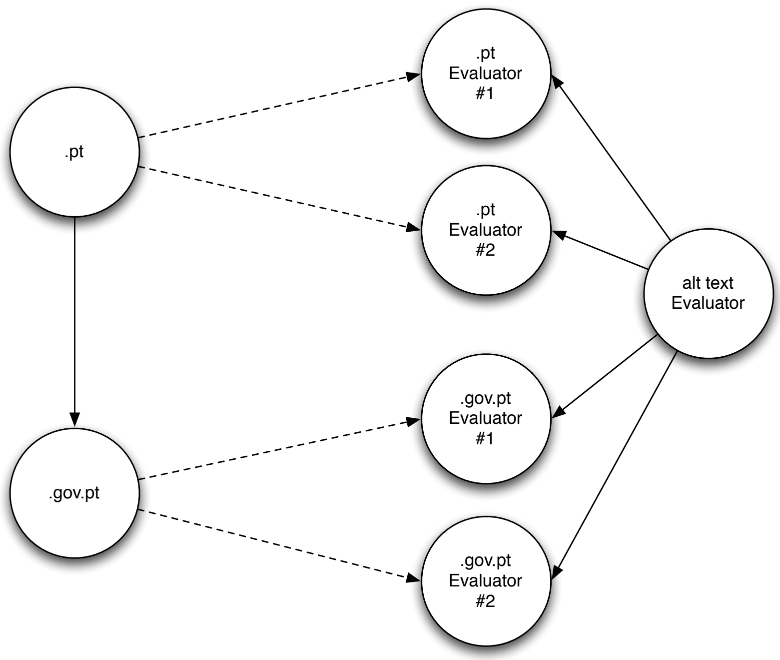


Figure 5.6: Mixing Evaluators for Web Accessibility

5.3 Engineering Web Accessibility Evaluation at the Large

The previous Section discussed how *technological* and *topological* distributions of Web accessibility evaluators can be combined, in order to form a mesh of evaluators that can support large scale accessibility evaluations of Web pages. While this model provides the ground for evaluation practices, it is still necessary to understand how evaluation results are to be interpreted at such scales, in order to understand how microscopic properties of accessibility (e.g., the evaluation result of a single Web page) lead the way to macroscopic properties.

To materialise this vision, we focused on putting the Evaluation And Report Language (EARL) at the core of a Web Science lifecycle, as depicted in Figure 5.7.

This lifecycle has been defined as an expansion of EARL concepts with Linked Data, which allowed us to define a Web of accessibility evaluation data. Each step in this lifecyle provides a key point in its definition:

- *Issues*: it all starts with the motivation for this thesis, i.e., the fact that there is no overall vision of Web accessibility, with regards to how accessibility guidelines are implemented throughout the Web;

- *Idea*: the focal concept behind this thesis, which is the merge between *Linked Data* and *Web accessibility evaluation* to afford a large-scale analysis of the state of Web accessibility;

- *Tech*: the definition of distributed models for Web accessibility evaluation, and its application to existing evaluation practices and standards, e.g., extending EARL to cope with proposed models;

- *Social*: since having the distributed technology ready is not enough, *social* factors must be taken into account to implement the distributed model across TLDs and descendent domains. Consequently, there must be a co-operative push by accessibility stakeholders such as accessibility experts, governments, etc. for the success of the proposed model;

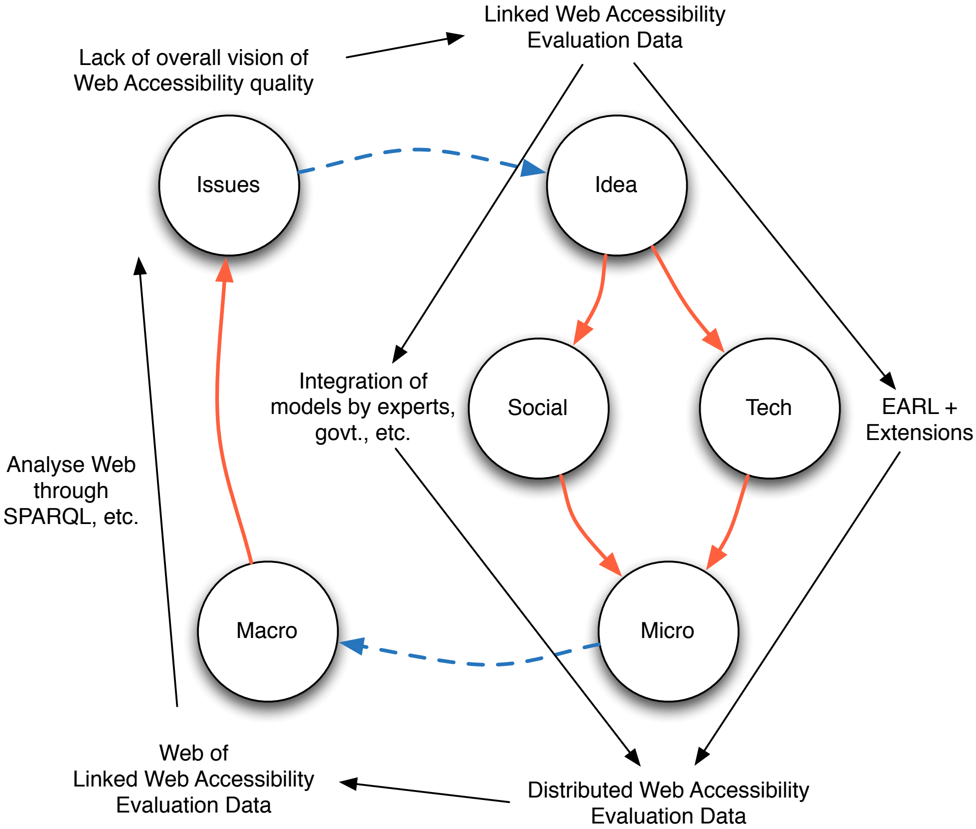


Figure 5.7: Applying EARL and Linked Data into the Web Science Lifecycle

- *Micro*: having *technical* and *social* factors defined, we reach the goal of providing means for obtaining distributed data about Web accessibility evaluation;

- *Macro*: escalating the *micro* stage into the entire Web, we finally have a Web of *Linked Data* conveying information about accessibility evaluation of, potentially, every Web page.

At this point, having a *Web of Linked Web Accessibility Evaluation Data* provides a solid ground for studying Web accessibility at the large, e.g., through the use of SPARQL. This factor will help completing the proposed lifecycle, by providing the tools to discover and understand what issues are in the way of having an accessible Web for everyone. In the next Sections we present a set of extensions to EARL for supporting this lifecycle, as well as how to extract useful information of the state of Web accessibility at the large.

5.3.1 Describing the Accessibility Quality of the Web

As discussed in Chapter 3, the Web Accessibility Initiative at W3C defined the EARL language as the way to describe the accessibility of a Web page. With EARL, through a set of constructs on top of the RDF framework, it is possible to represent the several factors of accessibility evaluation results, e.g., made with automated Web accessibility evaluation software: *someone checks a resource* and *resource fails a test*. In our distributed model for Web accessibility evaluation, each *technological* evaluator is responsible for the verification of a (sub)set of accessibility guidelines, for specific Web technologies, which results in a machine-processeable evauation report described with EARL. Figure 5.8 presents this relationship between *technological* evaluators and EARL.

EARL, at its core, provides sufficient constructs to describe the accessibility quality of a Web page, or a set of Web pages viewed as a single entity (i.e., through the combination of the `earl:TestSubject` and `dct:hasPart` constructs[1]). However, the linking structure that binds Web pages to each other is not represented

[1]The presented namespace prefixes follow EARL's conventions as defined in http://www.w3.org/TR/EARL10-Schema/#conventions

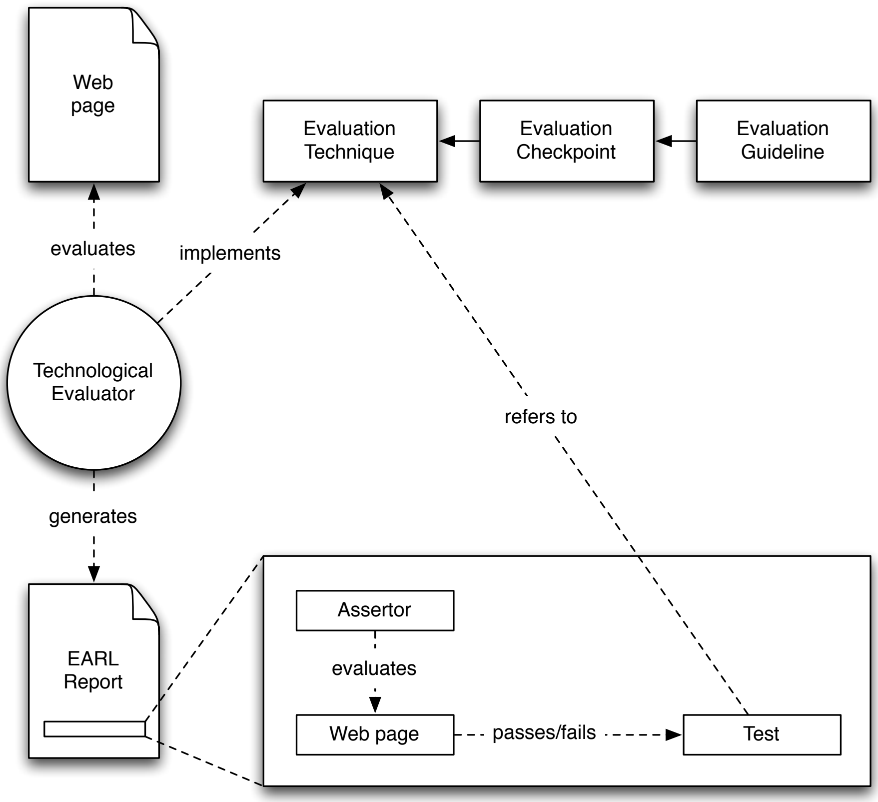


Figure 5.8: Relationship between *Technological* Evaluators and EARL

in EARL. To afford Linked Data principles, EARL documents must be connected in the same way the Web pages they represent are connected (Figure 5.9).

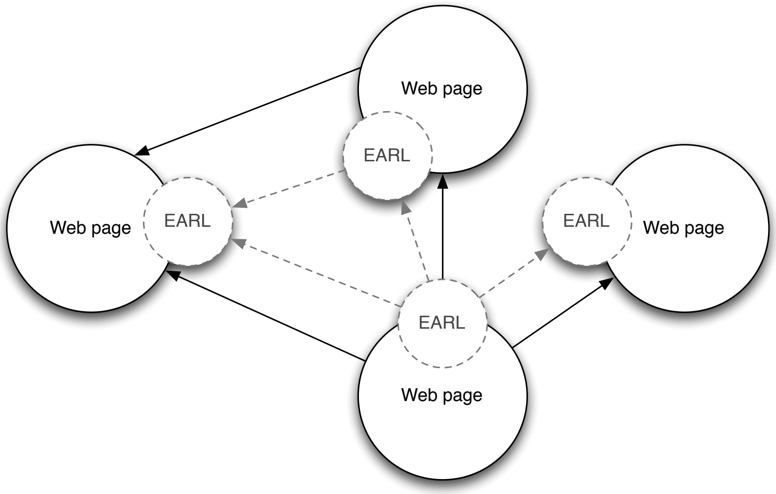


Figure 5.9: Overlaying EARL and Linked Data on Web Topology

This approach also opens the way to the analysis of Web accessibility evaluation centred on mining *Linked Data ready* EARL documents, thus skipping the inspection of Web pages themselves, i.e., allowing for the direct crawling of EARL documents.

To accommodate this capability onto EARL, we devised a small RDF vocabulary to afford Linked Data concepts of connecting distributed data (in this case EARL documents), in the form of an RDF *Property* `qw:evaluationLinksTo`[1]. The full RDF triple to represent this information is completed by having the *Resource* and *Literal* parts correspond to EARL documents, for instance:

```
http://example.com/earl.rdf qw:evaluationLinksTo http://example2.com/earl.rdf
```

5.3.2 Expanding to Web Accessibility Metrics

EARL provides a set of constructs to describe the evaluation result of a test, such as `earl:passed`, `earl:failed`, `earl:cantTell`, amongst others. However,

[1]The `qw` namespace prefix maps to the QualWeb project URI http://hcim.di.fc.ul.pt/wiki/QualWeb#

this information provides a limited view of a Web page's overall accessibility quality, as thoroughly discussed in the State-of-the-Art Chapter. Since quantification metrics allow for a more precise analysis of Web accessibility quality, we defined another extension to EARL to allow the representation of Web accessibility metrics (and subsequent metric application to Web pages), as presented in Figure 5.10:

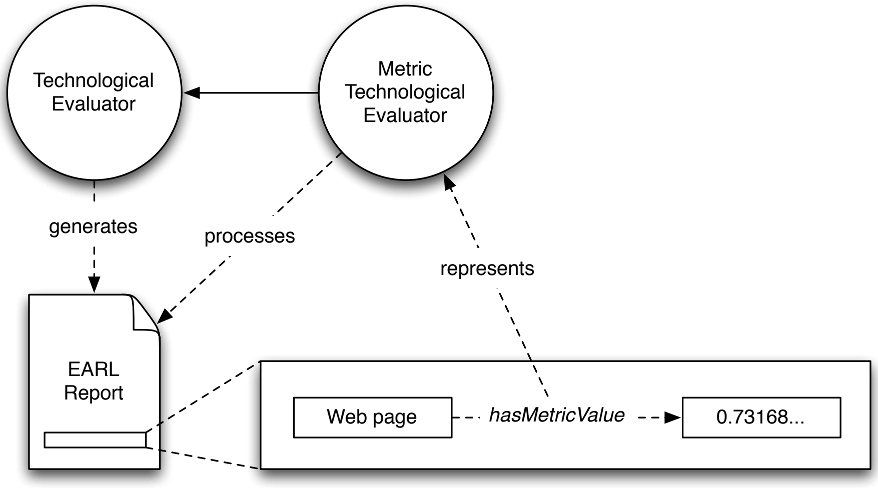


Figure 5.10: Extending EARL with Web Accessibility Metrics

This extension takes the form of an RDF *Property*, `qw:hasMetricValue`, which binds together a Web page *Resource* to a quantification *Literal* (e.g., *floating point number*). Since metrics differ from each other (Freire *et al.*, 2008), each metric has to extend `qw:hasMetricValue` accordingly, by means of the `rdfs:subPropertyOf` RDFS construct. Find below example usage:

- *UWEM Metric*:

  `http://example.com/page.html qw:hasUWEMMetricValue 0.75`

  assuming the RDF triple

  `qw:hasUWEMMetricValue rdfs:subPropertyOf qw:hasMetricValue;`

- *WCAG Compliance*:

  `http://example.com/page.html qw:hasWCAGComplianceMetricValue AA`

assuming the RDF triple
`qw:hasWCAGComplianceMetricValue rdfs:subPropertyOf qw:hasMetricValue.`

5.4 Querying Web Accessibility from Linked EARL Reports

Having accessibility evaluation reports detailed in EARL, linking to each other, opens the way for leveraging existing Semantic Web technologies to inspect and extract useful information about the state of accessibility on the Web. One of such technologies is SPARQL, one of the query languages to extract information from RDF data graphs. Thus, the challenge lays at formulating SPARQL queries that allow for knowledge discovery about accessibility of the Web. In the following Sections we detail some example queries in 3 domains: *Web pages*, *hyperlinks*, and *meta analysis*.

5.4.1 Web Page Accessibility Quality

The most direct analysis of *Linked Data* based on EARL reports is to assess the accessibility quality of Web pages. This includes checking the state of guideline compliance, analyse metrics' values, etc.

- *Which Web pages pass on WCAG 2.0 Guideline 1.1?*

```
SELECT ?page WHERE {
  ?assertion rdf:type earl:Assertion .
  ?assertion earl:subject ?page .
  ?assertion earl:test <http://www.w3.org/TR/WCAG20/#text-equiv> .
  ?assertion earl:result ?result .
  ?result earl:outcome earl:passed
}
```

Listing 5.1: SPARQL Query for Guideline Assessment

- *Which Web pages comply with WCAG 2.0 Level AA?*

```
SELECT ?page WHERE {
  ?assertion rdf:type earl:Assertion .
  ?assertion earl:subject ?page .
  ?assertion qw:hasWCAGComplianceMetricValue AA
```

```
}
```

Listing 5.2: SPARQL Query for WCAG Level Compliance

5.4.2 Hyperlinking Data Analysis

The previous examples present the basis of analysis of EARL reports. However, exploiting the linked structure of the Web can also leverage new findings about accessibility on the Web, as presented in the following SPARQL examples.

- *What Web pages linked from a specific Web page have a UWEM compliance value > 0.75?*

```
SELECT ?page WHERE {
  ?source qw:evaluationLinksTo ?page .

  ?srcassertion rdf:type earl:Assertion.
  ?srcassertion earl:subject ?source .

  ?assertion rdf:type earl:Assertion .
  ?assertion earl:subject ?page .
  ?assertion qw:hasUWEMMetricValue ?value .
  FILTER (?value > 0.75)
}
```

Listing 5.3: SPARQL Query for Linking Structure Based on UWEM Compliance

- *What is the Web's sub-graph with WAQM compliance value > 0.9?*

```
CONSTRUCT {
  ?page1 xlink:href ?page2
}
WHERE {
  ?page1 qw:evaluationLinksTo ?page2 .

  ?assertion rdf:type earl:Assertion .
  ?assertion earl:subject ?page1 .
  ?assertion qw:hasWAQMMetricValue ?value .

  ?assertion2 rdf:type earl:Assertion .
  ?assertion2 earl:subject ?page2 .
  ?assertion2 qw:hasWAQMMetricValue ?value2 .

  FILTER (?value > 0.9 && ?value2 > 0.9)
}
```

Listing 5.4: SPARQL Query for Extracting a Web Graph Based on WAQM Compliance

5.4.3 Evaluation Meta-Analysis

Finally, by having a rich semantic information about Web accessibility evaluation allows for the study not just of how it overlays the Web, but also how the evaluation process itself is devised.

- *Which evaluators pass WCAG 2.0 Guideline 1.1 for a specific Web page?*

```
SELECT ?evaluator WHERE {
  ?assertion rdf:type earl:Assertion .
  ?assertion earl:assertedBy ?evaluator .
  ?assertion earl:subject <http://example.com/test.html> .
  ?assertion earl:test <http://www.w3.org/TR/WCAG20/#text-equiv> .
  ?assertion earl:result ?result .
  ?result earl:outcome earl:passed
}
```

Listing 5.5: SPARQL Query for Evaluators Compliance

- *Which metrics return a value > 0.5 for a specific Web page?*

```
SELECT ?metric WHERE {
  ?assertion rdf:type earl:Assertion .
  ?assertion earl:subject <http://example.com/test.html> .
  ?assertion ?metric ?value .

  FILTER (?value > 0.5)
}
```

Listing 5.6: SPARQL Query for Metrics

5.5 Discussion

In the previous Sections we proposed a new model for Web accessibility evaluation, based on a *divide and conquer* approach to study Web accessibility at a large scale. By having this new model deployed, it will open the way to understanding what are the macroscopic forces of Web accessibility. However, to fulfill this vision, several assumptions and opportunities arise in the way of its implementation. In the next Sections we discuss some of these issues, including: *services*, *consistency*, *undecidability*, *trust*, *precision and recall*, and *authorship*.

5.5.1 Building Services on Top of Linked Web Accessibility Evaluation Data

Until now, several accessibility stakeholders involved in the process of developing accessibility evaluation and repair solutions – researchers, developers, accessibility experts, etc. – spend significant effort to devise and reimplement Web accessibility evaluation software. On the other hand, the high entry barrier for studying Web accessibility at the scale of the Web deters researchers from the goals of Web science. With a proper implementation of the model we proposed (i.e., *social* dimension within the Web science lifecycle), stakeholders' efforts can be shifted on building services and knowledge on top of *Linked Web Accessibility Evaluation Data*, thus aiming at answering questions on Web accessibility from a Web science perspective. The availability of Linked EARL reports can open the way to improvements in the different directions, including:

- *Accessibility Crawlers.* Just like for Web pages, the availability of a Web of Linked Data based on EARL reports has the potential to open the way for *accessibility crawlers*. We define these crawlers as software agents that mine EARL reports and follow the links between them, in order to mine Web accessibility properties from this information graph. This type of information can be used, e.g., to augment search engine ranking algorithms with the accessibility quality of each Web page implicit on accessibility metrics (Vigo *et al.*, 2009a).

- *Accessibility Observatories.* Another group of beneficiaries from our proposed approach concerns *observatories for Web accessibility*. While earlier initiatives to create these observatories have been made[1], their results and consequences are less than desired. The existence of accessibility observatories allow for a constant analysis of the state of accessibility quality of

[1]European Internet Accessibility Observatory: http://www.eiao.net/

the Web, data that can be used to devise new inclusion policies, guidelines, as well as education strategies. With the model we propose in this thesis, *topological* evaluators provide a solid framework for the creation of these observatories.

- *Reporting and Repair Tools.* The availability of EARL reports already opened the way for decoupling Web accessibility evaluation, reporting, and the presentation of this information to key stakeholders such as developers and accessibility experts. However, by having these reports linked together, a new opportunity rises for Linked Data consumption and visualisation tools tailored to Web accessibility evaluation reporting. This kind of tools can be leveraged by accessibility researchers, in order to help them grasping macroscopic properties of accessibility on the Web.

- *Assistive Technologies (ATs).* The main stakeholders on Web accessibility, i.e., people with disabilities, can also benefit from the availability of a Web of Linked Data comprising of EARL reports. Through advancements on assistive technologies' binding to the proposed distributed model, users can have augmented navigation capabilities between Web pages, e.g., through navigation scents based on Web page accessibility quality (Vigo *et al.*, 2009b).

5.5.2 The Undecidability of Automated Web Accessibility Evaluation

Not all evaluation techniques for Web accessibility are currently automatable (Velleman *et al.*, 2007), such as univocally deciding if the alternative text of an image conveys the appropriate information. While efforts are being made in this direction, e.g., through machine learning techniques (Olsen *et al.*, 2010), there is still a gap between what can be evaluated automatically and manually, as depicted in Figure 5.11.

However, in large-scale Web accessibility evaluations, automation is the key. It is suspected that evaluation *warnings* (automated inspections based on partially-undecidable evaluation techniques) are proportional to *errors* (machine-verifiable

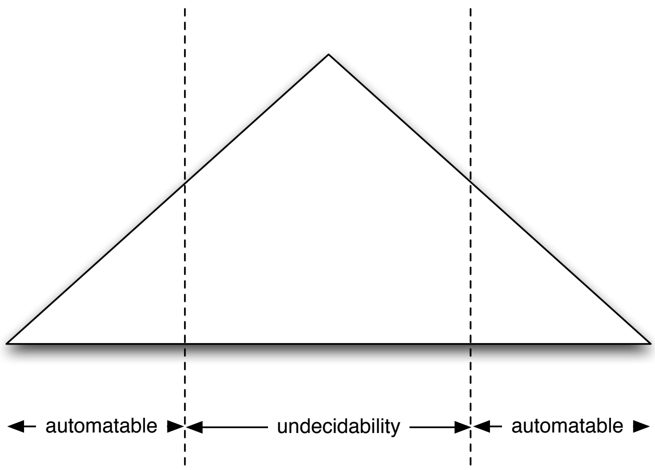


Figure 5.11: Undecidability Model for Automated Web Accessibility Evaluation

assertions of accessibility compliance) for each Web page (Vigo *et al.*, 2007). Nevertheless, the effective machine-based assertion of accessibility compliance will shed the light on several accessibility properties of the Web.

5.5.3 Web Accessibility and Content Production Detection

In Section 2 we discussed the WAI model and how it relates to evaluation and reporting through EARL. However, nowadays, there is a strong shift of the Web towards content production by end users, which raises more issues on the accessibility of Web pages. We extended both developer and user roles into other expected behaviours on the process of bringing accessibility to the Web, as presented in Figure 5.12.

This extension to the WAI Model is detailed as follows:

- Users are also shifting to *content production* roles for the Web, e.g., through blogs and wikis. Hence, the software they use to create this content must be accessible too, as defined by ATAG, and the produced content itself must comply with WCAG;
- Developers are, ultimately, responsible for *creating* assistive technologies and user agents that comply with UUAG, coming up with solutions to aid users with disabilities on interacting with the Web.

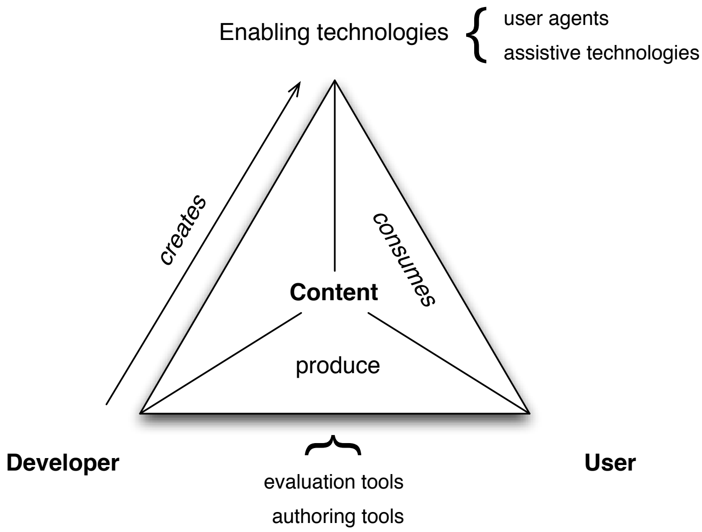


Figure 5.12: Reframing Components of the WAI Model

This shift raises concerns on understanding how accessibility on the Web is changing. In an earlier experimental study on the accessibility quality of Wikipedia, as detailed on Chapter 4, we learned that HTML template engines help stabilising the accessibility quality of Web sites, but the openness of Content Management Systems (CMS) fosters their instability. Therefore, it is important to detect which content production stakeholders (developers, designers, end users, etc.) are responsible for detected problems on the accessibility of a given Web page, in order to understand the broadness of applicability of ATAG and WCAG.

EARL currently provides a way to bind specific portions of evaluated content (e.g., an HTML element) to evaluation outcomes (Iglesias & Squillace, 2009), but the *production* nature of the content is still difficult to grasp (e.g., if it was delivered through a CMS). The detection of CMS-based versus hand-coded content has been tackled previously (Baeza-Yates *et al.*, 2007), as well as applying reverse-engineering techniques to detect Web application models (Bouillon *et al.*, 2002; Paganelli & Paterno, 2002). We believe that this type of techniques can help fostering the analysis of Web accessibility quality from a content production point of view.

5.6 Summary

This Chapter presented a novel approach to understand how the Web Science and Web Accessibility disciplines are orthogonal to each other, while having the same overarching goal of making the Web accessible (in *lato sensu*) to everyone. We defined a new model for Web accessibility evaluation based on a distributed architecture for Web accessibility evaluation software, to afford the Web science lifecycle. On top of this proposal, we presented a set of extensions to the EARL report language for Web accessibility evaluation that bind it to the goals of *Linked Data*, as well as to accessibility quality metrics. This allowed us to explore how to query a *Web of Linked Web Accessibility Evaluation Data*, in order to extract meaningful knowledge on the state of accessibility on the Web.

These proposals raised a set of questions and future challenges for the Web accessibility research field, including *services*, *consistency*, *undecidability*, *trust*, *precision and recall*, and *authorship*. We believe that, with the implementation of our proposed model, Web accessibility stakeholders will be able to grasp the macroscopic properties of the Web that influence Web accessibility, as well as the properties of accessibility on the Web.

Finally, the next Chapter discusses how this modelling framework can be engineered, specially from the implementation and deployment points-of-view.

Vision without implementation
is hallucination

Benjamin Franklin

Chapter 6

The QualWeb Evaluator

6.1 Introduction

The previous two Chapters have detailed experimental studies and models for Web accessibility evaluation at the large. Both discussed several challenges that face the exploration of the hypothesis raised by this thesis. This Chapter presents the most significant challenges of implementing a large-scale Web accessibility evaluator. These are presented in the context of the experimental studies, as well as how to achieve the goals proposed by the distributed Web accessibility evaluation model.

6.2 Implementation

This Section discusses the most significant aspects of implementing Web accessibility evaluation at the large. We begin by describing the core evaluation component – named *QualWeb* –, and afterwards present its application in *scaling* and *distribution* scenarios.

6.2.1 Accessibility Evaluator

Implementing automated Web accessibility evaluation, as described earlier in Chapters 2 and 3, requires the inspection of front-end Web technologies, such as HTML, CSS, or Javascript (and any combination of them). To meet the demands of large-scale evaluations, we defined a set of specific requirements, as presented next. Afterwards, we discuss some design and implementation issues that have been encountered on the QualWeb accessibility evaluator.

6.2.1.1 Requirements

Evaluating Web accessibility at the large (in the order of millions of Web pages) imposes several restrictions on how an evaluation component must be implemented, including:

- *Robustness*: the structure of HTML documents is widely variable. There is a high probability of encountering ill-formed documents that do not respect the proper structures defined in HTML document type definitions[1]. Consequently, all tasks that involve parsing and analysing HTML documents must be *robust* and be able to infer as many details as possible them;

- *Resources*: since the proposal of this thesis concerns scalability, it is necessary to minimize resource allocation as possible. This includes *memory*, *CPU*, and *network* resources. Thus, the computational cost for accessibility evaluation must be optimized (e.g., number of evaluations/second).

From the analysis of existing Web accessibility evaluators (as presented earlier in the State-of-the-Art Chapter), no one met these two overarching requirements. On *robustness*, most implementations leveraged existing XML-based parsers, which can easily break on ill-formed HTML. Furthermore, the HTML specification allows for document structures that are not XML-compatible (e.g., lack of quotes on element attributes, missing attribute values, and no enforcement on closing atomic elements). On the *resources* side, the minimisation of *network* resources dismissed completely the implementation of accessibility evaluation based on externally-available software (e.g., through a Web service), since these are not prepared for evaluating a large magnitude of Web pages. For the other resources, we found out that all evaluators – due to their interactive nature – require a significant amount of memory and CPU resources.

Consequently, we overcame these limitations by implementing a high-performant Web accessibility evaluator. Next, some design and implementation choices are further detailed.

6.2.1.2 Design

The evaluation of a Web page' accessibility quality is an operation that takes a given *input* – a Web page and its ancillary resources – and returns an *output* detailing the evaluation results. In the case of the *QualWeb* evaluator, it was divided into the following sub-operations:

1. *Parsing*: we used the NekoHTML parser as a starting point to process HTML documents, since it does not block the two requirements elicited

[1] DTD for HTML 4.01: http://www.w3.org/TR/html4/sgml/dtd.html

above: it handles ill-formed HTML without failing (*robustness*), and provides SAX-based APIs[1] to afford a stream-based processing of HTML documents, thus reducing memory and CPU footprints (i.e., better handling of *resources* for processing activities).

2. *HTML element iteration*: after setting up the generation of SAX events, we hooked Groovy's APIs for tree-based processing. This allowed us to write different accessibility evaluation techniques (based on WCAG) as native methods and having the benefits of SAX-based processing. We opted for a *depth-first visiting* algorithm to each HTML element in the document representation, and apply evaluation operations to each element, as follows:

```
html.depthFirst().each { node ->
  checkpoints.each { cp ->
    resultsList = []

    if (cp.applicable(node)) {
      cp.evaluateAndCalc(node)
    }

    resultsList << cp.getCurrent()
  }

  this.resultsPerNode << resultsList
}
```

Listing 6.1: Generic Depth First Evaluation

Here, in Listing 6.1, `html` represents a handle to the root element of the HTML document, upon which the *depth-first visiting* algorithm iterates. Each `node` in the tree is used as an input to be tested by all `checkpoints`. And, in each checkpoint, two sub-operations are performed as follows:

(a) *Applicability*: each HTML element is tested if it is a target for the current checkpoint being evaluated. For instance, in the case of checkpoint 1.1, testing for all `img` elements will yield a positive answer. This allows for keeping trace of how many HTML elements are valid targets for each checkpoint – to be used to compute accessibility metrics;

[1]SAX: Streaming API for XML

(b) *Evaluation*: if a specific checkpoint is applicable to the current `node`, then the appropriate evaluation procedure is executed.

Each checkpoint extends specific functionality provided by a common execution plan, as presented below in Listing 6.2:

```
public void evaluateAndCalc(node) {
  tested++

  this.current = evaluate(node)

  switch (this.current) {
    case RESULT_PASS:
      passed++
      break

    case RESULT_FAIL:
      failed++
      break

    case RESULT_WARN:
      warned++
      break
  }
}
```

Listing 6.2: Generic Checkpoint Evaluation

Afterwards, all evaluation results per checkpoint per node are stored for further processing. Next, a sample implementation of a WCAG checkpoint is presented.

6.2.1.3 Example WCAG Implementation

In Chapter 3 it was presented how WCAG compliance works – i.e., by properly encoding accessibility features into HTML and other Web front-end technologies – and how it can be exploited by software-based evaluators. Likewise, the QualWeb evaluator follows the same strategy of HTML inspection. Below is a fragment implementation of WCAG 1.0 Checkpoint 1.1, where HTML element applicability (Listing 6.3) and HTML element evaluation (Listing 6.4) are performed:

```
private boolean applicable_html_01(node) {
  return \
```

```
    (node.name() in ["img", "area", "applet"]) ||
    (node.name() == "input" && node.attributes().type == "image")
}
```

Listing 6.3: WCAG 1.0 Checkpoint 1.1 Applicability Criterion

```
public int evaluate(node) {
  if (applicable_html_01(node) && (node.attributes().alt == null))
    return RESULT_FAIL

  if (applicable_html_06(node))
    return RESULT_FAIL

  if (applicable_html_02(node))
    return RESULT_WARN

  return RESULT_PASS
}
```

Listing 6.4: WCAG 1.0 Checkpoint 1.1 Evaluation

The application of this checkpoint to all nodes requires $O(n)$ computational time, where n represents the number of HTML elements in the HTML document. Furthermore, by being applied in a stream of elements (provided by the SAX API), the required memory cost is constant (i.e., one element at a time).

For the cases where checkpoints require the comparison between different HTML elements (such as Checkpoint 12.4 where `label` elements should be attached to corresponding `form` elements), the use of caching and dynamic programming algorithms allowed the maintenance of computational time complexity and memory consumption at the same level.

The implementation of this accessibility evaluation component was self-contained in order to be used in different evaluation scenarios. Consequently, different interfacing capabilities were introduced into the component, including:

- *Embedded*: the QualWeb evaluator has been compiled as a *.jar* component, which can be programmatically leveraged in Java-compatible execution environments;
- *Service*: a thin *Web Service* wrapper was created, which allows for the deployment of the QualWeb evaluator in any computer, and to be interfaced with different programming languages.

As discussed above, the resulting outcome of the evaluation process is a set of accumulators that describe how many times a given checkpoint was evaluated, and their corresponding evaluation results. Optionally, these results are complemented with the exact HTML location where each evaluation has been applied.

This information forms the basis of the evaluation *output*, to be further processed according to different evaluation scenarios. For instance, in the case of the large scale experimental study, these results were aggregated in a compact form (in the *CSV* format), due to sheer amount of Web pages that were evaluated. For other scenarios, an EARL-based result is provided. Nevertheless, the QualWeb evaluator provides report templating mechanisms that allow its simple extension to different serialisation formats.

6.2.2 Scaling the Evaluation

The experimental study presented in Chapter 4 applied the QualWeb evaluation component through its *embedded* interface. This study was conducted on the *Hadoop* distributed computing platform built on top of a server cluster. Since Hadoop provides means to parallelise computational activities, we leveraged this capability to perform assessment tasks with several instances of the evaluator. As the QualWeb evaluator does not introduce dependencies between evaluated Web pages, this process can be scaled up horizontally. The computational workflow devised for the study follows:

- *Splitting*: the evaluated HTML document collection was split across all servers in the cluster;
- *Distributing*: the Hadoop platform was configured so that each server had each core per CPU running one Hadoop task. Here, each document subcollection is constantly fed to each Hadoop task, consecutively;
- *Processing*: each task instantiates and/or resets an instance of the QualWeb evaluator, and provides it with an HTML document. The evaluating process is triggered;
- *Storing*: finally, the resulting outcome from each evaluation is stored sequentially on each server.

6.2.3 Distributed Evaluation

In order to engineer the distributed evaluation proposal described earlier in Chapter 5, two challenges arise: *integration* and *delegation*, corresponding to the core functionalities of *technological* and *topologic* evaluators, respectively. Next, each challenge is further described.

6.2.3.1 Integration

The availability of different evaluators tailored to specific *technologies* introduces the problem of integrating partial evaluations. For instance, to fully assess the accessibility of a given Web page, *technological* evaluators for HTML, CSS, and Javascript (amongst other front-end technologies) have to process different parts of the content accordingly. The combination of their outcomes forms the final evaluation report, generated by a *composite* evaluator.

In the case of the proposed distributed evaluation model, reports are specified in the EARL document format. Consequently, the combination of individual reports generated by *atomic technological* evaluators must be a valid EARL document. Since EARL is an application of RDF, this task was trivially achieved with the aid of RDF manipulation libraries, as exemplified by the small code sample presented below in Listing 6.5 (using the Jena RDF Framework):

```
def combineEARLReports(reportsList, finalReportFilename) {
  def finalModel = ModelFactory.createOntologyModel()

  reportsList.each { report ->
    reportModel = ModelFactory.createOntologyModel()
    reportModel.read(report)

    finalModel.addSubModel(reportModel)
  }

  finalModel.validate()
  finalModel.prepare()

  finalModel.write(new FileOutputStream(finalReportFilename))
}
```

Listing 6.5: Combined EARL Report Generation

This aggregation assumes the availability of each EARL report within the evaluation process. The *QualWeb* evaluator can act as a *composite* evaluator and apply this small method to aggregate its evaluation results with a foreign CSS-centric accessibility evaluator (such as verifying if all font sizes are specified in *em* units). Since the proposed model is distributed, one cannot expect to have all technological evaluators available within the same execution space. Consequently, these must be externally available. This is achieved through a *delegation* process, as detailed next.

6.2.3.2 Delegation

The other challenge concerns the implementation of *delegating* the evaluation process, according to a *topological* order. There is already a large-scale Internet deployment of such type of systems: DNS – Domain Name System (Mockapetris, 1987). However, since Linked Data operates on top of the Web, we emulated the DNS system behaviour of *name caches* on top of HTTP in a distributed way. The implementation strategy is as follows: given a URL referencing a Web page that is to be evaluated, e.g., http://www.di.fc.ul.pt/, the distributed model tries to find the most specific subdomain that has an available accessibility evaluator that serves as the authority for all pages falling under its domain and subdomains. For this to happen, each domain must know what subdomains under its control have an evaluator available.

Consider the aforementioned URL, and the following *name caches*: $pt \rightarrow \{ul, up\}$, $ul \rightarrow \{fc, fml\}$. Here, each token represents a URL that can be queried for an accessibility evaluation (e.g., such as the Web service provided by *QualWeb*). *pt*, *ul.pt*, *fc.ul.pt*, *fml.ul.pt*, and *up.pt* domains have accessibility evaluators available to be used in their subdomains.

To implement such mechanism, we leveraged the HTTP response *303 See Other* status code (which is also commonly used for Linked Data implementations). When a given evaluation service is asked for an accessibility evaluation for a URL, it emits a *303* response to a subdomain, when appropriate (according to the rules specified above).

Using the example above, for simplicity reasons, we will assume that an accessibility query service knows the locations of all evaluators for Top Level Domains (TLDs), such as *.pt*, evaluator URLs are under the form http://evaluator.*domain*, and → represents an *HTTP 303 Redirect*. A possible execution run for this algorithm is:

1. http://evaluator.pt/ → http://evaluator.ul.pt

2. http://evaluator.ul.pt → http://evaluator.fc.ul.pt

When reaching the final URL, http://evaluator.fc.ul.pt, the evaluation process is triggered, and an EARL report is emitted as the final evaluation result, according to the *composite technological evaluator* rules described above. This simple implementation allows for a distributed model for accessibility evaluation that puts the onus of specificity and accountability on domain and subdomain owners. Consequently, it can be freely scaled, just like other distributed technologies on the Web (which allows for leveraging existing Web scalability technologies, such as proxies and caches).

6.3 Summary

This Chapter presented the major points of the implementation for the proposed Web accessibility evaluation methodologies. Since most accessibility evaluators are devised for small scale experiments or to aid developers producing accessible Web content, there is little need to cope with scalability and performance issues. However, for the type of studies presented in this thesis, these issues became relevant.

We described how the *QualWeb* evaluator was developed, and how scalability and performance was taken into account in this process. On top of this evaluator, we discussed some scenarios where it has been applied, including *embedded* in distributed computing platforms, as well as *services* for distributed accessibility evaluation.

Finally, we detailed the roadmap taken to bring to life the proposed distributed Web accessibility evaluation model, where *atomic* and *composite* evaluators are distributed in a *topological* mesh. This work allows for the deployment of distributed Web accessibility evaluators which open the way to large-scale assessments and establishment of accessibility observatories. Additional scenarios for this evaluation model are further detailed in the next Chapter as *Open Challenges*.

The great tragedy of Science –
the slaying of a beautiful
hypothesis by an ugly fact.

Thomas Henry Huxley

Chapter 7

Conclusions and Open Challenges

This thesis presented a comprehension over different aspects that entail a Web-scale characterisation of Web accessibility quality. It is know that most Web developers and designers do not have knowledge on accessibility. This results on the known fact that, overall, the Web is not accessible for people with disabilities. However, several of the factors that lead to this – and their repercussion as well – are still a black box, since Web accessibility is studied at a microscopic level, i.e., at the Web page or Web site scope.

The work conducted through this dissertation started with a moderately sized sample of Web pages that was experimentally evaluated with regards to its accessibility quality. This has lead to understanding that HTML template mechanisms help stabilising the accessibility quality of the Web. The study also uncovered the variability on accessibility quality between different Web sites.

On top of these results, we conducted a large-scale analysis of the accessibility quality of a collection of Web pages provided by the Portuguese Web Archive. Through a formulation of accessibility quantitative metrics, this study revealed the profound differences imposed by the interpretation of results of automated Web accessibility evaluators by Web developers and designers. Another revelation concerns the correlation between Web page complexity (in terms of HTML elements) and its accessibility quality: simpler, smaller Web pages tend to be more accessible. Finally, template mechanisms were observed at large-scale as well.

The effort required for building Web accessibility evaluation workbenches for large-scale analysis is high, as described in Chapters 4 and 6. In order to attain the long-term vision of understanding the state of accessibility of the Web in its entire size – e.g., by establishing Web accessibility observatories – this thesis also proposes a distributed model for Web accessibility evaluation, which uncovered also some problems on the accessibility evaluation process.

In the next Sections, these results are revisited in the context of the initial research hypothesis. Finally, this dissertation presents the major challenges for Web accessibility evaluation uncovered by this work.

7.1 Revisiting the Hypothesis

This dissertation started with a problem statement: despite the fact that it is technically possible to create Web pages that are accessible to users with disabilities, most of the Web is inaccessible somehow. Four main factors have been described as the main reasons for this: *Users and the Inaccessible Web*, *Developers' Unawareness*, *End-user Content Production*, and *Large-scale Accessibility Evaluation*. These problems have been verified - up to a certain point - in small scales (i.e., microscopic), but not at a large scale (i.e., macroscopic). The link between the microscopic and the macroscopic state of the Web's accessibility has been unknown. This problem served as the basis for the hypothesis that conducted the work detailed on this dissertation, as presented again below:

> *Web accessibility evaluation at large scale will leverage new knowledge about the Web itself, its development and adequate mechanisms to raise proper awareness of designers and developers towards accessibility.*

In general terms, this hypothesis was verified with the experimental studies described in Chapter 4. The large-scale experiment conducted over the Portuguese Web Archive collections has shown that Web accessibility does follow a pattern, in that accessibility quality generally increases when page complexity (measured in terms of HTML elements) decreases. The breakdown of accessibility quantification metrics into distinction between the possible outcome of an evaluation *warning* (i.e., its falsehood or truthiness) has shown its impact on accessibility awareness, since it is known that accessibility warnings are often dismissed by developers and designers.

Our first experimental study confirmed that the usage of HTML templates helps stabilising accessibility quality. But when analysing the linking structure of the Web pages, it was clear that accessibility was not a concern on whether the link should exist or not. This confirms the influence of end-users on produced content and corresponding accessibility. When applying the same concepts to the large-scale study, the same pattern emerges: there is a tendency for template-based Web pages to be more accessible than others.

The large-scale experimental study has helped emerging new knowledge about the accessibility state of the Web. However, it is yet to be shown how this can be correlated with other structural properties of the Web. Consequently, this thesis presented a model for distributed accessibility evaluation – and a reference prototype implementation – that can be leveraged for these studies.

7.2 Open Challenges

This thesis is not, by any measure, a complete study on the application of Web science methodologies into Web accessibility practices (and vice-versa). The discussion Sections of Chapters 4 and 5 uncovered some of these open challenges, including:

- *Hyperlinking*: linking is one of the principal ways people connect and share information on the Web. It has been shown that this feature has a profound impact on the shape of the Web, which is recurrently exploited by search engines. While this thesis provides some insights onto hyperlinking and accessibility, there is still a huge challenge on understanding how accessibility relates with hyperlinking. For instance, studies by Vigo *et al.* (2009a) show that search results (and corresponding ranking algorithms) can be tweaked through accessibility metrics. However, the large-scale deployment and awareness of such practices might impose a reshaping of the Web and its linking structure (e.g., people will link to more accessible content, in order to get better ranking position);

- *Delivery Contexts*: nowadays the Web is being accessed in different devices, such as mobile phones, tablets, TVs, amongst others. This poses several challenges to different stakeholders. For *developers* and *designers*, they must cope with the limitations and capabilities of the platforms, such as screen size, available input methods, connectivity, and computational power. This leads to the creation of tailored Web sites that offer device-specific user interfaces according to its capabilities. One one hand, as discussed by Harper *et al.* (2006), these limitations can foster a more accessible Web, since *end-users* share some limitations with the devices. It is yet still a very nascent research field, so that these phenomena can be studied entirely on the Web. On the other hand, devices might impose new accessibility problems to users. The impact of this device shift is yet unknown;

- *Dynamics*: the current trends of the Web are also shifting into more dynamic content, where HTML documents change through interactions from users. With techniques such as WAI-ARIA (Craig & Cooper, 2010) it is possible to enrich Web applications so that they become accessible for users with disabilities. However, the extent upon which these techniques are applied, as well as how they can be detected, is becoming an active field within the accessibility research community;

- *Scalability*: our study provided clues on how accessibility evaluation practices can be scaled up to the size of the Web, both through proper implementation and distributed evaluation practices. However, it is impossible to predict the effect and impact of a large-scale deployment of continuing monitorisation of accessibility through the distributed accessibility evaluation model presented.

- *Automatability*: our experimental study has shown the deep difference imposed by the interpretation of evaluation *warnings*. Ideally, as detailed in Chapter 5, evaluation should be 100% automatic, so that it becomes objective and fully scaled to the size of the Web. However, several challenges are still in the way of understanding if a given content is accessible or not. Research fields such as image processing and analysis, information retrieval, reverse engineering, application modelling, or even intelligent agents, will provide advancements that can be throughly applied to reach this goal.

- *Eventual Consistency*: the architecture of the Web defines a set of HTTP methods that can be performed over Web resources (Fielding, 2000), such as GET, POST, etc. While some of these methods (e.g., GET, PUT) are idempotent, the inherent stateless characteristic of the HTTP protocol can lead to eventual consistency problems (Vogels, 2008) when accessing resources at different time instants. This leads to the fact that two GET operations to a given Web resource can return different content. Furthermore, it is known that several Web sites repurpose their content in consecutive requests for the same Web resource, thus making the Web permeable to the *observer effect* – each request for a resource returns a different content.

These intrinsic characteristics of the Web impact the proposed distributed model for Web accessibility evaluation and, consequently, on the services than can be built on top of it. We framed this problem into two scenarios:

1. *Observing.* User agents request Web page x to a Web server, which returns its current representation x_r. Assistive Technologies acting on top of the user agent request a Linked EARL report to a *technological* evaluator, which triggers a new evaluation of the x resource. However, this request affects the resource, which leads to the Web server returning a new representation x_{r+1}.
2. *Caching.* An *accessibility crawler* requests an EARL report for a given Web page x to a TLD *topological* evaluator. Due to computational constraints, this evaluator frequently crawls the Web and performs Web accessibility evaluations in an offline mode, storing the corresponding linked EARL reports. This evaluator returns an EARL report to the crawler that refers to an older representation x_{r-1} of the requested resource's current representation x_r.

While both cases differ in what actions trigger inconsistency, they lead to the same result: a mismatch between a Web page and its corresponding EARL report. This includes not just a different accessibility evaluation result set, but also a different linkage between EARL reports (i.e., *link rotten*).

- *Truthfulness*: we also delved into the fabric of accessibility evaluation results. The landscape of evaluators is high – when analysing automated software – and even higher when including humans in the process. Consequently, it is still unknown how truthful are the results that emerge from evaluations. While several implementations share the same common knowledge on what can be accessible or not (including the *QualWeb* evaluator), the lack of a platform for comparing evaluation results is still lacking. This will provide the basis for scientifically determining the accuracy of an accessibility evaluation, just like what is being done for information retrieval algorithms.

- *Trust and Accountability*: Trust on the Web is a core issue tackled by Web science, which defines three major challenges in this domain (Golbeck, 2008): *trust management* (policy), *computing trust* (quantification), and *applications using trust* (leveraging) – the definition of trust policies will lead to its computation, which can then be leveraged by miscellanea applications. Orthogonally, trust is also a matter of scope: *content*, *services*, and *people*. The use of Web accessibility evaluation software implies a certain model of trust for the results it presents: can an EARL document be trusted on its assertions? Can *technological* evaluators be trusted by *topological* evaluators? Can a researcher trust on those who developed evaluators?

  Furthermore, the impact of trust in Web accessibility evaluation goes beyond understanding how it is shaped on the Web. As explained earlier, several countries define their accessibility laws based on guidelines such as WCAG. Therefore, when leveraging our proposed model for distributed Web accessibility evaluation, trust must be taken into account in contexts such as governmental Web accessibility observatories, etc. The introduction of trust in this model is, therefore, of the uttermost importance for its successful implementation.

  With the ever increasing improvement of laws across the planet that foster the adequacy of Web content to people with disabilities, the relevance of law abidance increases. While content production is being increasingly monitored, incorrect evaluations are still hindering the real accessibility quality of the Web. The definition of an accountability process, akin to quality assurance and monitoring, is still lacking. This issue becomes even more relevant when scaled up to the size of the Web (not just governmental Web sites, as often is perceived).

References

ABASCAL, J., ARRUE, M., FAJARDO, I., GARAY, N. & TOMÁS, J. (2004). The use of guidelines to automatically verify web accessibility. *Univers. Access Inf. Soc.*, **3**, 71–79. 90

ABOU-ZAHRA, S. & SQUILLACE, M. (2009). Evaluation and report language (EARL) 1.0 schema. Last call WD, W3C, http://www.w3.org/TR/2009/WD-EARL10-Schema-20091029/. 48, 53

AIZPURUA, A., ARRUE, M., VIGO, M. & ABASCAL, J. (2009a). Exploring automatic css accessibility evaluation. In M. Gaedke, M. Grossniklaus & O. Díaz, eds., *Web Engineering*, vol. 5648 of *Lecture Notes in Computer Science*, 16–29, Springer Berlin / Heidelberg. 92

AIZPURUA, A., ARRUE, M., VIGO, M. & ABASCAL, J. (2009b). Transition of accessibility evaluation tools to new standards. In *W4A '09: Proceedings of the 2009 International Cross-Disciplinary Conference on Web Accessibililty (W4A)*, 36–44, ACM, New York, NY, USA. 90

APACHE FOUNDATION (2010). Hadoop. http://hadoop.apache.org/. 72

BAEZA-YATES, R., CASTILLO, C. & EFTHIMIADIS, E.N. (2007). Characterization of national web domains. *ACM Trans. Internet Technol.*, **7**, 9. 34, 82, 90, 109

BARRETO, A. (2008). Visual impairments. In *Web Accessibility*, 3–13, Springer, London, United Kingdom. 25

Beckett, D. & McBride, B. (2004). RDF/XML Syntax Specification (Revised). Tech. rep., World Wide Web Consortium (W3C), last accessed on April 9th, 2010, from http://www.w3.org/TR/rdf-syntax-grammar. 21

Berners-Lee, T. (2007). The two magics of web science. Keynote speech at the 16th International World Wide Web Conference, http://www.w3.org/2007/Talks/0509-www-keynote-tbl/. xvii, 32

Berners-Lee, T. (2009). Linked-data - design issues. W3C Design Issue Document, last accessed on April 9th, 2010, from http://www.w3.org/DesignIssues/LinkedData.html. 23

Berners-Lee, T., Cailleau, R., Groff, J. & Pollerman, B. (1992). World Wide Web: The Information Universe. *Electronic Networking: Research, Applications and Policy*. 3

Berners-Lee, T., Fielding, R. & Masinter, L. (2005). Rfc 3986: Uniform resource identifier (uri): Generic syntax. Tech. rep., Internet Engineering Task Force, last accessed on April 9th, 2010, from http://www.ietf.org/rfc/rfc3986.txt. 20

Berners-lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A. & Sheets, D. (2006). Tabulator: Exploring and analyzing linked data on the semantic web. In *In Procedings of the 3rd International Semantic Web User Interaction Workshop (SWUI06)*. 22

Berners-Lee, T., Hall, W., Hendler, J.A., O'Hara, K., Shadbolt, N. & Weitzner, D.J. (2006). A Framework for Web Science. *Found. Trends Web Sci.*, **1**, 1–130. 4, 31

Bigham, J.P. (2007). Increasing web accessibility by automatically judging alternative text quality. In *IUI '07: Proceedings of the 12th international conference on Intelligent user interfaces*, 349–352, ACM, New York, NY, USA. 90, 92

Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R. & Hellmann, S. (2009). Dbpedia - a crystallization point for the web of data. *Web Semant.*, **7**, 154–165. 38

Boley, H., Hallmark, G., Kifer, M., Paschke, A., Polleres, A. & Reynolds, D. (2010). Rif core dialect. W3C recommendation, W3C, http://www.w3.org/TR/rif-core/. 22

Bos, B., Çelik, T., Hickson, I. & Lie, H.W. (2007). Cascading Style Sheets, level 2 revision 1 – CSS 2.1 Specification. Tech. rep., World Wide Web Consortium (W3C), http://www.w3.org/TR/CSS21. 5

Bouillon, L., Vanderdonckt, J. & Eisenstein, J. (2002). Model-Based Approaches to Reengineering Web Pages. In *TAMODIA '02: Proceedings of the First International Workshop on Task Models and Diagrams for User Interface Design*, 86–95, INFOREC Publishing House Bucharest. 109

Brajnik, G. (2009). Validity and reliability of web accessibility guidelines. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*, Assets '09, 131–138, ACM, New York, NY, USA. 45

Brajnik, G., Mulas, A. & Pitton, C. (2007). Effects of sampling methods on web accessibility evaluations. In *Assets '07: Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, 59–66, ACM, New York, NY, USA. 90

Brajnik, G., Yesilada, Y. & Harper, S. (2010). Testability and validity of wcag 2.0: the expertise effect. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*, ASSETS '10, 43–50, ACM, New York, NY, USA. 45

Brickley, D. & Guha, R. (2004). Rdf vocabulary description language 1.0: Rdf schema. W3C recommendation, W3C, http://www.w3.org/TR/rdf-schema/. 22

Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. In *WWW7: Proceedings of the seventh international conference on World Wide Web 7*, 107–117, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands. 31, 82

Burner, M. & Kahle, B. (1996). Arc File Format. Tech. rep., The Internet Archive, http://www.archive.org/web/researcher/ArcFileFormat.php. 71

Bush, V. (1945). As we may think. *Atlantic Monthly*, **176**, 101–108. 3

Bühler, C., Heck, H., Perlick, O., Nietzio, A. & Ulltveit-Moe, N. (2006). Interpreting Results from Large Scale Automatic Evaluation of Web Accessibility. In *Computers Helping People with Special Needs*, 184–191, Springer. 52, 53

Caldwell, B., Cooper, M., Chisholm, W., Reid, L.G. & Vanderheiden, G. (2008). Web Content Accessibility Guidelines 2.0. W3C Recommendation, World Wide Web Consortium (W3C), http://www.w3.org/TR/WCAG20/. 6, 39, 41

Cardoso, N. & Silva, M.J. (2010). A gir architecture with semantic-flavored query reformulation. In *6th Workshop on Geographic Information Retrieval*, ACM, Zurich, Switzerland. 38

Carmody, S., Gross, W., Nelson, T., Rice, D. & van Dam, A. (1969). A Hypertext Editing System for the /360. Tech. rep., Center for Computer & Information Sciences, Brown University. 3

Chisholm, W., Vanderheiden, G. & Jacobs, I. (1999). Web Content Accessibility Guidelines 1.0. W3C Recommendation, World Wide Web Consortium (W3C), http://www.w3.org/TR/WCAG10/. 71

Chisholm, W., Vanderheiden, G. & Jacobs, I. (2000). Techniques for Web Content Accessibility Guidelines 1.0. W3C Note, World Wide Web Consortium (W3C), http://www.w3.org/TR/WAI-WEBCONTENT-TECHS/. 40

Christensen, N. (2005). Preserving the Bits of the Danish Web. In *5th International Web Archiving Workshop (IWAW05)*. 69

Chung, S., Shiowattana, D., Dmitriev, P. & Chan, S. (2009). The web of nations. In *WWW '09: Proceedings of the 18th international conference on World wide web*, 1159–1160, ACM, New York, NY, USA. 70

Clark, J. & DeRose, S.J. (1999). Xml path language (xpath) version 1.0. W3C Recommendation, World Wide Web Consortium (W3C), http://www.w3.org/TR/xpath. 72

Craig, J. & Cooper, M. (2010). Accessible rich internet applications (wai-aria) 1.0. W3C working draft, W3C, http://www.w3.org/TR/wai-aria/. 128

de Kunder, M. (2010). World Wide Web size. Last accessed on April 6th, 2010, from http://www.worldwidewebsize.com/. 3

Dean, J. & Ghemawat, S. (2004). Mapreduce: Simplified data processing on large cluster. *OSDI*, http://labs.google.com/papers/mapreduce-osdi2004.pdf. 72

Dill, S., Kumar, R., Mccurley, K.S., Rajagopalan, S., Sivakumar, D. & Tomkins, A. (2002). Self-similarity in the web. *ACM Trans. Internet Technol.*, **2**, 205–223. 33, 81, 90

Engelbart, D.C. & English, W.K. (1968). A Research Center for Augmenting Human Intellect. In *Computer-supported cooperative work: a book of readings*, 395–410. 3

Fielding, R., Gettys, J., Mogul, J.C., Frystyk, H., Masinter, L., Leach, P. & Berners-Lee, T. (1998). Hypertext Transfer Protocol – HTTP/1.1. Tech. Rep. Internet RFC 2616, IETF, http://www.ietf.org/rfc/rfc2616.txt. 5, 19

Fielding, R.T. (2000). *Architectural Styles and the Design of Network-based Software Architectures*. Ph.D. thesis, University of California, Irvine, Irvine, California. 128

Ford, K., Richards, J., Allan, J. & Spellman, J. (2009). User agent accessibility guidelines (UAAG) 2.0. W3C working draft, W3C, http://www.w3.org/TR/2009/WD-UAAG20-20090723/. 39

Freire, A.P., Fortes, R.P.M., Turine, M.A.S. & Paiva, D.M.B. (2008). An evaluation of web accessibility metrics based on their attributes. In *SIGDOC '08: Proceedings of the 26th annual ACM international conference on Design of communication*, 73–80, ACM, New York, NY, USA. 102

Golbeck, J. (2008). Trust on the World Wide Web: A Survey. *Found. Trends Web Sci.*, **1**, 131–197. 130

Gomes, D. & Silva, M.J. (2006). Modelling information persistence on the web. In *ICWE '06: Proceedings of the 6th international conference on Web engineering*, 193–200, ACM Press, New York, NY, USA. 69

Gomes, D., Nogueira, A., Miranda, J. & Costa, M. (2008). Introducing the portuguese web archive initiative. In *Proceedings of the 8th International Web Archiving Workshop*, Aarhus, Denmark.

Halpin, H. (2009). A query-driven characterization of linked data. In *Linked Data on the Web (LDOW2009)*, CEUR-WS. 38

Harper, S. & Yesilada, Y. (2008). *Web Accessibility*. Springer, London, United Kingdom. 5, 7

Harper, S., Yesilada, Y. & Goble, C. (2006). Building the mobile web: rediscovering accessibility?: W4A - International Cross-Disciplinary Workshop on Web Accessibility workshop report – 2006. *SIGACCESS Access. Comput.*, 21–32. 26, 127

Hendler, J., Shadbolt, N., Hall, W., Berners-Lee, T. & Weitzner, D. (2008). Web science: an interdisciplinary approach to understanding the web. *Commun. ACM*, **51**, 60–69. 31

Henry, S.L. (2005). Essential components of web accessibility. Last accessed on June 22nd, 2010, from http://www.w3.org/WAI/intro/components.php. 38

Henzinger, M.R., Heydon, A., Mitzenmacher, M. & Najork, M. (2000). On near-uniform url sampling. In *Proceedings of the 9th international World*

Wide Web conference on Computer networks : the international journal of computer and telecommunications netowrking, 295–308, North-Holland Publishing Co., Amsterdam, The Netherlands, The Netherlands. 70

Heydon, A. & Najork, M. (1999). Mercator: A scalable, extensible web crawler. *World Wide Web*, **2**, 219–229. 70

Hu, M., Lim, E.P., Sun, A., Lauw, H.W. & Vuong, B.Q. (2007). Measuring article quality in wikipedia: models and evaluation. In *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, 243–252, ACM, New York, NY, USA. 60

Höhrmann, B., Hégaret, P.L. & Pixley, T. (2006). Document Object Model (DOM) Level 3 Events Specification. Tech. rep., World Wide Web Consortium, http://www.w3.org/TR/DOM-Level-3-Events. 5

Iglesias, C. & Squillace, M. (2009). Pointer methods in rdf 1.0. W3C working draft, W3C, http://www.w3.org/TR/Pointers-in-RDF10/. 109

ISO (2003). ISO/TS 16071:2003: Ergonomics of human-system interaction – Guidance on accessibility for human-computer interfaces. Tech. rep., International Organization for Standardization. 26

ISO (2005). ISO 9241: Ergonomics of Human System Interaction. Tech. rep., International Organization for Standardization. 23

Jacobs, I. & Walsh, N. (2004). Architecture of the World Wide Web, Volume One. W3C Recommendation, World Wide Web Consortium (W3C), last accessed on April 9th, 2010, from http://www.w3.org/TR/webarch/. 20

Kahle, B. (2002). The Internet Archive. *RLG Diginews*, **6**. 69

Keates, S. & Clarkson, P.J. (2003). Countering design exclusion through inclusive design. In *CUU '03: Proceedings of the 2003 conference on Universal usability*, 69–76, ACM, New York, NY, USA. 26

Kellar, M., Watters, C. & Shepherd, M. (2006). A Goal-based Classification of Web Information Tasks. In *Proceedings of the 69th Annual Meeting of the American Society for Information Science and Technology*, ASIS&T, Austin, TX, USA. 19

Koster, M. (1994). A standard for robot exclusion. Tech. rep., http://www.robotstxt.org/wc/robots.html. 70

Li, D., Browne, G. & Wetherbe, J. (2006). Why Do Internet Users Stick with a Specific Web Site? A Relationship Perspective. *IJEC*, **10**, 105–141. 4

Lie, H.W. (2005). CSS3 module: Cascading and inheritance. W3C Working Draft, World Wide Web Consortium (W3C), http://www.w3.org/TR/css3-cascade/.

Lopes, R. & Carriço, L. (2008). The impact of accessibility assessment in macro scale universal usability studies of the web. In *W4A '08: Proceedings of the 2008 international cross-disciplinary conference on Web accessibility (W4A)*, 5–14, ACM, New York, NY, USA. 82

Lopes, R., Gomes, D. & Carriço, L. (2010a). Web not for all: A large scale study of web accessibility. In *W4A: 7th ACM International Cross-Disciplinary Conference on Web Accessibility*, ACM, Raleigh, North Carolina, USA. 93

Lopes, R., Isacker, K.V. & Carriço, L. (2010b). Redefining assumptions: Accessibility and its stakeholders. In *The 12th International Conference on Computers Helping People with Special Needs (ICCHP)*, Springer. 6

Lunn, D., Bechhofer, S. & Harper, S. (2008). The sadie transcoding platform. In *W4A '08: Proceedings of the 2008 international cross-disciplinary conference on Web accessibility (W4A)*, 128–129, ACM, New York, NY, USA. 48

Manola, F. & Miller, E. (2004). Rdf primer. Tech. rep., World Wide Web Consortium (W3C), last accessed on April 9th, 2010, from http://www.w3.org/TR/rdf-primer/. 22

MIRANDA, J. & GOMES, D. (2009). An Updated Portrait of the Portuguese Web. In *14th Portuguese Conference on Artificial Intelligence (EPIA 2009)*, Aveiro, Portugal. 70, 71

MOCKAPETRIS, P. (1987). *RFC 1034 Domain Names - Concepts and Facilities*. Internet Engineering Task Force, http://tools.ietf.org/html/rfc1034. 93, 120

MOHR, G., KIMPTON, M., STACK, M. & RANITOVIC, I. (2004). Introduction to heritrix, an archival quality web crawler. In *4th International Web Archiving Workshop (IWAW04)*. 71

NELSON, T.H. (1980). *Literary Machines*. Mindful Press. 3

NORMAN, D.A. (1998). *The Invisible Computer: Why good products can fail, the personal computer is so complex, and information appliances are the solution*. MIT Press. 4

NTOULAS, A., CHO, J. & OLSTON, C. (2004). What's new on the web?: the evolution of the web from a search engine perspective. In *Proceedings of the 13th international conference on World Wide Web*, 1–12, ACM Press. 69

OLSEN, M., SNAPRUD, M. & NIETZIO, A. (2010). Automatic checking of alternative texts on web pages. *Computers Helping People with Special Needs*. 80, 92, 107

PAGANELLI, L. & PATERNO, F. (2002). Automatic Reconstruction of the Underlying Interaction Design of Web Applications. In *SEKE '02: Proceedings of the 14th international conference on Software engineering and knowledge engineering*, 439–445, ACM, New York, NY, USA. 109

PEMBERTON, S., AUSTIN, D., AXELSSON, J., ÇELIK, T., DOMINIAK, D., ELENBAAS, H., EPPERSON, B., ISHIKAWA, M., MATSUI, S., MCCARRON, S., NAVARRO, A., PERUVEMBA, S., RELYEA, R., SCHNITZENBAUMER, S. & STARK, P. (2002). XHTML 1.0 The Extensible HyperText Markup Language (Second Edition). Tech. rep., World Wide Web Consortium (W3C), http://www.w3.org/TR/xhtml1. 5, 20

Philips, M. (2003). PANDORA, Australia's Web Archive, and the Digital Archiving System that Supports it. *DigiCULT.info*, 24. 69

Preece, J., Sharp, H., Benyon, D., Holland, S. & Carey, T. (1994). *Human Computer Interaction*. Addison-Wesley, Reading, Massachusetts. 23

Probst, K., Johnson, B., Mukherjee, A., van der Poel, E. & Xiao, L. (2009). A proposal for making AJAX crawlable. Tech. rep., Google, http://googlewebmastercentral.blogspot.com/2009/10/proposal-for-making-ajax-crawlable.html.

Prud'hommeaux, E. & Seaborne, A. (2008). SPARQL Query Language for RDF. Tech. rep., World Wide Web Consortium, last accessed on April 9th, 2010, from http://www.w3.org/TR/rdf-sparql-query/. 22

Schreiber, G. & Dean, M. (2004). OWL Web Ontology Language Reference. W3C Recommendation, World Wide Web Consortium, last accessed on April 9th, 2010, from http://www.w3.org/TR/owl-ref/. 22

Shadbolt, N. (2008). Web science research roadmap. Last accessed on April 7th, 2010, from http://webscience.org/research/roadmap.html. 31

Shneiderman, B. (2000). Universal Usability. *Commun. ACM*, **43**, 84–91. 5

Shneiderman, B. (2007). Web Science: A Provocative Invitation to Computer Science. *Commun. ACM*, **50**, 25–27. 5

Sloan, D., Heath, A., Hamilton, F., Kelly, B., Petrie, H. & Phipps, L. (2006). Contextual web accessibility - maximizing the benefit of accessibility guidelines. In *W4A '06: Proceedings of the 2006 international cross-disciplinary workshop on Web accessibility (W4A)*, 121–131, ACM, New York, NY, USA. 73

Sullivan, T. & Matson, R. (2000). Barriers to use: usability and content accessibility on the web's most popular sites. In *CUU '00: Proceedings on the 2000 conference on Universal Usability*, 139–144, ACM, New York, NY, USA. 51, 62, 73

Treviranus, J., Richards, J. & Spellman, J. (2009). Authoring tool accessibility guidelines (ATAG) 2.0. W3C working draft, W3C, http://www.w3.org/TR/2009/WD-ATAG20-20091029/. 39

US Congress (1998). Section 508 Amendment to the Rehabilitation Act of 1973. http://www.section508.gov/. 26

Velasco, C.A. & Koch, J. (2009). Evaluation and report language (EARL) 1.0 guide. W3C working draft, W3C, http://www.w3.org/TR/2009/WD-EARL10-Guide-20091029/. 53

Velleman, E., Meerveld, C., Strobbe, C., Koch, J., Velasco, C.A., Snaprud, M. & Nietzio, A. (2007). Unified Web Evaluation Methodology (UWEM 1.2). 46, 51, 61, 72, 90, 107

Vigo, M., Arrue, M., Brajnik, G., Lomuscio, R. & Abascal, J. (2007). Quantitative metrics for measuring web accessibility. In *W4A '07: Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A)*, 99–107, ACM, New York, NY, USA. 52, 79, 92, 108

Vigo, M., Arrue, M. & Abascal, J. (2009a). Enriching information retrieval results with web accessibility measurement. *Journal of Web Engineering*, **8**, 3–24. 106, 127

Vigo, M., Leporini, B. & Paternò, F. (2009b). Enriching web information scent for blind users. In *Assets '09: Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*, 123–130, ACM, New York, NY, USA. 107

Vogels, W. (2008). Eventually consistent. *Queue*, **6**, 14–19. 128

Voss, J. (2005). Measuring wikipedia. In *Proc. ISSI 2005*, Stockholm, Sweden. 60

W3C (2006). Policies relating to web accessibility. http://www.w3.org/WAI/Policy/. 48

WEINREICH, H., OBENDORF, H., HERDER, E. & MAYER, M. (2006). Off the beaten tracks: exploring three aspects of web navigation. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, 133–142, ACM, New York, NY, USA. 34

WIKIPEDIA CONTRIBUTORS (2011). Wikipedia:Dead external links. *Wikipedia, The Free Encyclopedia*, accessed June 27, 2011 from http://en.wikipedia.org/wiki/Wikipedia:Dead_external_links. 68

WILKINSON, D.M. & HUBERMAN, B.A. (2007). Cooperation and quality in wikipedia. In *WikiSym '07: Proceedings of the 2007 international symposium on Wikis*, 157–164, ACM, New York, NY, USA. 60

YESILADA, Y., BRAJNIK, G. & HARPER, S. (2009). How much does expertise matter?: a barrier walkthrough study with experts and non-experts. In *Assets '09: Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*, 203–210, ACM, New York, NY, USA. 45, 81

ZENG, X. (2004). *Evaluation and Enhancement of Web Content Accessibility for Persons with Disabilities*. Ph.D. thesis, University of Pittsburgh. 51