Study Plan

This cover page must be completed and returned together with the study plan to The International Doctoral School of Technology and Science, the Faculties of Engineering, Science and Medicine, Niels Jernes Vej 10, 9220 Aalborg Øst, Denmark

PhD Programme: Computer Science and Engineering

Project title: eGovernment Monitoring with focus on Machine Perception

Name of the Ph.D. student: Morten Goodwin Olsen

Education: Master of Science Information and Communication Technology

Institution: University of Agder

Supervisor: Odd Mikael Holmesland Snaprud

Co-supervisor (optional): Christian S. Jensen

Department: Computer Science

Date of enrolment: 2008-10-15

Expected date of completion: 2011-09-01

Signature

2009-10-28		
Date	Ph.D. student	
		CHRISTIAN S. JENSEN
Study plan approved		
2008-10-22		MIKAEL SNAPRUD
Date	Supervisor	Printed name
30-10-08		KRISTIAN BLEEN
Date	Head of Department	Printed name
29-10-08		TORBEN BACH PEDERSEN
Date	Head of Ph.D. Programme	Printed name

Ph.D. Study Plan for Morten Goodwin Olsen

.

Version:	1.1
Author(s):	Morten Goodwin Olsen
Date:	October 30, 2008
Dissemination Level:	Public
Status:	Final

This document consists of 35 pages including this cover.

Abstract

This document outlines the Ph.D. study plan for Morten Goodwin Olsen on benchmarking eGovernment web services. The research is part of the eGovMon project which will evaluate and measure public web sites for accessibility, transparency, efficiency and impact.

The most significant part of the research will be creating methods for the measurements of all four areas of eGovMon. These metrics will consist of (1) Fully automatable tests (2) Heuristic tests and (3) Manual tests.

The heuristic tests will include so-called machine learning algorithms,which, based on some input or training data, will be able to perform tests, that are currently carried out manually.

Results from these evaluations will contribute to the awareness and improvement accessibility, transparency, efficiency and impact in public web sites.

Additionally, the learning algorithms which will be developed are expected to be a significant contribution in the field of web-mining.

Version Control

Ver.	Status	Date	Change	Author
0.1	DRAFT	2008-07-30	Initial Draft	Morten Goodwin Olsen
0.2	DRAFT	2008-07-31	Integrated comments from Mikael Snaprud. Initial draft of	Morten Goodwin Olsen
0.3	DRAFT	2008-08-29	Updated to fit new template as suggested by Christian Jensen.	Morten Goodwin Olsen
0.4	DRAFT	2008-10-04	Update of scientific content, budget, agreement, abstract goals, methods, courses.	Morten Goodwin Olsen
0.5	DRAFT	2008-10-04	Update according to comments from Christian Jensen and Mikael Snaprud	Morten Goodwin Olsen
0.6	DRAFT	2008-10-04	Wrote state-of-the art. Updated papers, dissemination external collaboration, more.	Morten Goodwin Olsen
0.7	DRAFT	2008-10-09	Fixed spelling and writing	Morten Goodwin Olsen
0.8	DRAFT	2008-10-09	Updated according to comments from Mikael Snaprud	Morten Goodwin Olsen
0.9	Release Candidate	2008-10-13	Updated according to comments from Torben Bach Pedersen	Morten Goodwin Olsen
1.0	Final	2008-10-28	Updated according to comments from Mikael Snaprud	Morten Goodwin Olsen
1.1	Final	2008-10-21	Updated according to comments from Torben Bach Pedersen. Added cover page	Morten Goodwin Olsen

Table 1: Version Control

Contents

List of Tables

List of Figures

1 Short Summary

The Ph.D. is part of the eGovMon [Sna08a] project where we plan automatically benchmark eGovernment services for four eGovernment indicators **Accessibility**, **Transparency**, **Efficiency** and **Impact**

The expected scientific output from this Ph.D. is divided into two main categories:

- An specification and implementation of an observatory which automatically benchmarks web sites according to accessibility, transparency, efficiency and impact. Several of the tests part of this observatory will be so-called learning algorithms − algorithms which, based on some input or training data, will be able to perform tests which typically are done manually.
- Empirical results of accessibility, transparency, efficiency and impact, both retrieved automatically using the observatory implementation and results from used surveys.

Several one-off studies within these topics have been conducted the past few years. Such studies have traditionally been conducted manually which could be both costly and biased.

In contrast, in my Ph.D. I plan to contribute to an automatic assessment that can easily be repeated at regular intervals with a significantly reduced cost compared to manual evaluations. Additonally, such an approach will remove any bias that is part of manual evaluations.

In this Ph.D. the most significant scientific evaluations will be carried out as a fully automatic empirical study automatically. This is intended to be supplemented by manual surveys done by the municipalities.

For measuring online eGovernment services web-mining techniques will be used, including the use of so-called machine learning algorithms. The main idea is for these algorithms to automatically learn what is transparent, efficient, accessible and has an impact based on some input data. This data could be user interaction or existing manual surveys. By doing this, the algorithms benefit directly from the manually organized data.

As an example, an open post list on public web sites containing content of mail interaction between the public institution and its users is considered transparent. Thus, for measuring transparency, it would be beneficial to detect if such list exists within a web site. There is no standard way of presenting post lists. Despite of this, any human who is familiar with the concept and understands the language, can easily detect if such a list is present.

Using machine learning algorithms will make it possible to detect significant web site features, and some of their properties, such as an open post list. Based on input data (so-called training data) provided for the algorithm, it could for example automatically learn which words are normally present in web pages presenting post lists and which words are normally not. Furthermore, the algorithm could, with a given perception, claim (classify) that a post list exists in a web site - removing or drastically reducing the need for human evaluation.

Another example as part of detecting efficiency within a public government agency could be based on e-mail interaction within the agency. Information such as who interacts with whom could be extracted. Furthermore, a graph of the interaction of the communication can be created and statistical data of the interaction can be extracted such as who are key persons within the agency, (finding hub nodes in clustering). This approach could be a significant input for experts determining the efficiency of the public agency.

In this Ph.D. I plan to evaluate, develop and apply learning algorithms for all the above four mentioned areas of eGovernment measurement.

2 Scientific Content

2.1 State-of-the-art

This section contains state of the art-description related to the Ph.D. This section is divided into the four areas: accessibility, transparency, efficiency and impact. Additionally, I suggest an architecture for an eGovMon Observatory evaluations where all four areas are four measuring all for areas including interacting with the user.

2.1.1 Web Accessibility

The most widely recognized resource when it comes to web accessibility is the Web Content Accessibility Guidelines (WCAG) 1.0 [Wor99] by the Web Accessibility Initiative (WAI),World Wide Web Consortium (W3C). These guidelines have become the de-facto standard for what is an accessible web site. A significant criticism of these guidelines have been regarding the fact the guidelines can not be applied as tests for web sites.

However, WCAG 2.0 [Wor08], which is scheduled for release in the end of 2008, is intended to be more testable. It contains both success criteria and common mistakes which could be applied to any given web page or web site in order to determine if it is accessible or not. Despite of this, even if all success criteria are met for a web site - it is no guarantee that the page in fact is accessible. Additionally, if none of the common mistakes are implemented, there is not guarantee that the web page or web site is accessible.

Furthermore, most of the success criteria and common mistakes require manual testing and can not directly be implemented as part of an automatic tool.

Web accessibility has received a lot of attention these last years. Most significantly, it is an important goal of the European i2010 strategy [Eur05]. It is clear that to know if this goal is reached, the accessibility of European web sites needs to be measured. However, the actual measurement of web accessibility is a challenging task which have mostly been conducted as manual studies.

Several studies on eAccessibility have been conducted over the past few years, most of these studies have been manual one-off studies. These are studies that have been conducted once where the web sites have been evaluated manually. An example of such a study which have received a lot of publicity is Measuring

Progress of eAccessibility in Europe (MeAC) [CKM07]. The MeAC assessment covers a wide variety of ICT products ranging from TV and telephony to computer hardware and software. The accessibility of public and private sector web sites is also part of the survey. For this study, in the parts regarding web accessibility, there are some clear limitations. For example, only 25 pages from each web site has been selected for evaluation. This is despite the fact that most web sites contain a substantially larger number of pages than 25 [OUMN+08]. This means, for this study, only a very small percentage of each web site has contributed to the web accessibility results.

Other studies which have received a lot of attention in the media is the report on eAccessibility of public sector services in the European Union by the UK cabinet office 2005 [UK 05] and United Nations Global Audit of Web Accessibility 2006 [Nom06]. Both studies include a survey of to what degree web sites conform to web accessibility standards and a ranking of web accessibility of each country. As different methodologies have been used in these studies, the results are not directly comparable with eachother.

Furthermore, several studies from individual European countries also exists based on the WCAG guidelines. Although the same guidelines are used, a range of different evaluation methodologies and scoring schemes are deployed across the member states, which makes it hard to compare the results between the states. To deal with this problem the Unified Web Evaluation Methodology (UWEM) [Web07, NSV08, NUMGO07] has been developed by the Web Accessibility Benchmarking (WAB) cluster. UWEM includes a complete methodology for evaluating and presenting accessibility of web sites. The methodology includes both automatic and manual testing. Only 20% of the UWEM tests are defined as automatable. This means that automatic accessibility testing according to UWEM can only present an indication of the accessibility of web sites, not claim that a web site is accessible. However, UWEM makes it possible to evaluate web sites within Europe using the same scheme, which further makes it possible to compare web sites evaluation between countries.

A fully automatic implementation of UWEM is the European Internet Accessibility Observatory (EIAO) [Sna08b, SOA06, UMOP+08]. EIAO has implemented the full UWEM methodology (sampling, evaluation, presentation) using the automatable tests. Even though EIAO only includes the automatable test (20% of all tests in UWEM) results showing a correlation between the automatic EIAO results from previous studies, such as MeAC [BHN+08], has been presented. It should be noted that this comparison has only been carried out using the disitribution of all results. There exists no study comparing the automatic EIAO results with manual evaluation on an individual web site or web page level.

All the automatable tests in UWEM and EIAO are fully deterministic tests which has some limitations. As an example, one of the automatable UWEM tests verifies that there is an alt attribute (alternative text) for an <img> element. However, the validity of such alternative texts is not part of the fully automatable UWEM tests. This means, for a web site to confirm to an automatic implementation of UWEM (such as EIAO), any alternative text is sufficient.

UWEM and EIAO presents a score of how accessible the evaluated web site is as:

$$\text{Accessibility Score} = \frac{\text{Number of detected barriers}}{\text{Total numbers of applied tests}}. \quad (1)$$

.

2.1.2 Transparency

Benchmarking transparency is a challenging task compared to accessibility as there is there exists no well defined and accepted guidelines stating what makes an online web service, such as a municipality web site, transparent.

Despite of this, there exists several surveys within the topic. As an example, Transparency International publishes an annual survey on corruption. Each year it focuses on one significant topic - in 2008 this was on corruption in the water sector [Tra08]. The annual corruption report also includes a survey on state-of-the-art research within the field of detecting and measuring corruption. An example of such is a survey on the confrontation of bribes in correlation with household income. Similar measurements would be very challenging, if not impossible, to extract from the web. Because of this, even though the topic is related to eGovMon, how the data is collected is far from how we could collect it from the web. In contrast, some parts of the methodology, such as scoring scheme, may be applicable to eGovMon.

Creating algorithms for measuring transparency is essential to the Ph.D. An interesting presentation of such results where done in the [Pav06], a paper on public procurement market transparency indicators. In this paper they surveyed the number of open biddings in the public Czech market, from which they defined a transparency score as following:

$$\text{transparency} = 100 * \frac{\text{Contracts awarded in an open bidding procedure}}{\text{Total volume of public procurement}} \quad (2)$$

The higher percentage of contracts awarded openly, the higher the transparency metric. It is noticeable that this score is rather similar to the UWEM score (see equation 1).

It is clear that for an end user it is beneficial to present all data in a clear and consistent manner. One way of doing this would be to present both accessibility results and transparency results in a similar way. Equation 2 and 1 clearly shows that this is possible.

So far, when it comes to transparency, I have only presented studies using off-line surveys and have not addressed transparency in an online matter.

Some indications of transparency available from the web has already been defined. For example [Ber08] describes five such five transparency indicators:

- **Document transparency**: The possibility of examining documents used as background for political decisions.
- **Benchmarking transparency**: The possibility to evaluate decision making including access plans, annual reports etc.

- **Meeting transparency**: The possibility to following governmental meetings (availability of proceedings and similar).
- **Decision makers transparency.** The possibility to know decisions made by and opinions of elected officials.
- **Disclosure transparency**: The possibility to retrieve information that has not published publically such as the possibility to ask questions regarding official documents and proceedings.

Indications of all of the above transparency indicators could be extracted directly from the web and/or interaction with public government agencies. This could be done ether automatically, semi-automatically or manually.

As an example, for document transparency, it is possible to check if the available documents contain political decisions or not. This could be achieved by using text classification algorithms ether completely automatically or semi-automatically [1].

Some benchmarking transparency of web sites exists. An example of this is the Dutch tool The Web Guidelines quality model ([Ove08]) which performs benchmarking according 125 tests, including transparency, organized into seven categories:

- **Standards**: e.g. is the web site organized according to the national web standards.
- **Transparency to the public**: e.g. is the budget of the government agency publicly available.
- **Organizing**: e.g. is the web site organized in such a way that it is quick and easy to find contact information.
- **Plans**: e.g. is plans of disaster relief available on the web sites.
- **Services**: e.g. is it possible to submit an environmental complaint.
- **Personalized Services**: e.g. is it possible for citizens to sign up for regular news letters.
- **Engagement and participation**: e.g. is it possible to discuss administrative or policy related items on the web site.
- **Accessibility**: e.g. is the webrichtlijnen-button available (web site has been automatically tests).

This tool, and the corresponding methodology, is mostly manual. The automatic evaluation is limited to link checking, HTML validation and checking of web site structure. The definitions of what is and what is not transparent from a web site, and is clearly useful for the algorithms to be created in this Ph.D.

[1] Note that such an approach, ether manual or automatic, can never be 100% accurate as it is possible to prevent publishing documents online that should have been published.

Another paper on measuring transparency is [Osi08] which focuses on measuring transparency of Web 2.0[2]. This paper claims that the increasing use and reuse of services which can been seen in Web 2.0 sites, such as the use of RSS and open APIs, should be a significant part of transparency measurements. According to the findings in this paper, transparency should be the key-factor in measuring any eGovernment services using Web 2.0 technology.

2.1.3 Efficiency

There exists some studies on measuring government which also includes efficiency, such as Understanding and Measuring eGovernment: International Benchmarking Studies [Hee06]. In this paper efficiency is seen as a subset of impact. I.e. one important component to have impact of an eGovernment service, is to have it efficient.

Some parts of the benchmarking intended to be done by eGovMon will be self-reporting. However, this paper claim that self-reporting is rather useless when the information will be publicly reported. According to the claim, the government agencies have a tendency to be bias whenever reporting good and bad actions when this information will be publically available. From this we can claim that self-reporting alone should not be a significant part of the eGovMon measurements, but could rather work as input/training data to some learning algorithms.

Another surveys which includes efficiency is [Mil08]. Here efficiency is defined as "impact for money" in contrast to effectiveness which is seen as quality of the existing services (e.g. search quality on a web site). Efficiency is seen as the key-element in any government service and the dilemma of getting most out of tax payers money is the most crucial issue, which again is very much related to impact. Performing actual measurements efficiency as it is defined here, is rather challenging to do ether automatically or manually - as we can not expect to get information such as the cost for each service.

On the other hand, effectiveness, as it is defined in [Mil08], should be possible to measure using learning algorithms. As an example, in an effective web site it should be easy to search a web site. This means that any user of the web site should be able to find what they are looking by searching for and intuitively understand the search results. Such an algorithm could have searched for some predefined documents, such as building regulation, and automatically find if it is available in the search results. Furthermore, detecting if the data is understandable could done by a classified algorithm.

2.1.4 Impact

Impact is possibly the most challenging of the four areas of eGovMon to measure automatically. In this area there exists very little indications that could be extracted automatically from the web. However, as shown in chapter 2.1.3,

[2]Web 2.0 is a high level definition of the emerging trends in the world wide web such as increased interactivity and interconnectivity [Wik08].

efficiency and impact is often seen as the same issue: an efficient service will have a good impact, and any inefficient service will not have a good impact.

Most available research on impact includes surveys sent out to users. Based on the reply, the impact of a product change etc. is measured. To the best of my knowledge, there exists no research on how to measure impact automatically from web sites, as it could be done with accessibility and transparency.

Some automatic ways of measuring impact could be to simply extract information of the use of the online services such as counting the number of hits on a particular web page. We can assume that a published service in for example a municipality, would have a lot of impact on the inhabitants if it gets a lot of hits, and very little impact if only few people visit the page. However, it is clear that this alone cannot be an indicator of impact.

Some studies exists such as [The03b] containing information about impact assessment for voluntary organizations who want to assess their impact, and [The03a] which identifies ways to demonstrate the impact of a voluntary organization. The impact of public eGovernment services is not part of these studies. In contrast they are intended as tools for organizations who want to show that their work matters. Despite of this, what is considered "good impact" and how this could be measured by a voluntary organization may be applicable for eGovMon.

2.1.5 eGovMon Architecture

The eGovMon observatory will build upon the already existing European Internet Accessibility Observatory (EIAO [Sna08b]). For details on the existing EIAO architecture, I refer the reader to [UMOP+08]. A proposed architecture where I expand the implementation by having four metrics for can be seen in figure 1.

The main work flow of the eGovMon observatory will be to be as following.

1. The URL repository is seeded with URLs of web sites to be evaluated.
2. The crawler downloads web pages from the web sites.
3. The sampler selects web pages randomly for evaluation from the downloaded web sites. This approach is a statistically sound near uniform random sampling. Each page is sent to all four metrics.
4. Automatic evaluation:
   (a) The accessibility metrics evaluated the web page for accessibility.
   (b) The transparency metrics evaluated the web page for transparency.
   (c) The efficiency metrics evaluated the web page for efficiency.
   (d) The impact metrics evaluated the web page for impact.
5. Manual evaluations:

   (a) Surveys are sent to different municipalities and public agencies. The results are incorporated together with the automatic results.
   (b) Experts evaluate selected web pages with tests that can only be done manually.

   Both of the manual above evaluations could work as input/training data to the learning algorithms. Each manual evaluation will increase the accuracy of the algorithms.

6. Results are stored in temporary RDF databases.
7. The ETL (extract transform load) extracts the results from the temporary RDF databases and inserted these into the Data Warehouse.
8. The evaluated results are presentable to users through online reports.

It should be noted that, in contrast to EIAO, the users will have a direct connection to the metrics. This will consists of:

- **Manual** surveys performed by the municipalities.
- Input, such as **training data**, for the learning algorithms.
- **Expert evaluation** ether as fully manual tests or as part of semi-automatic tests.

2.2 Project Objectives

The objective of eGovMon is to create a methodology and implement an observatory for measuring Accessibility, Transparency, Efficiency and Impact.

In this Ph.D. most of the focus will be on creating algorithms for these measurements. Many of these will consist of learning algorithms such as pattern recognition algorithms. In contrast to having only formal deterministic tests, using learning algorithms will make it possible to apply larger test sets including tests which are traditionally done manually.

2.3 Key methods

2.3.1 Method

Several research methods applicable in the Ph.D. The key methods in this Ph.D. can be organized as following.

1. **Literature review.**

   Reviewing for state-of-the-art literature including what is accessible, transparent, efficient and has an impact. This should include if is some of the areas have been benchmarked automatically previously and if so, how.

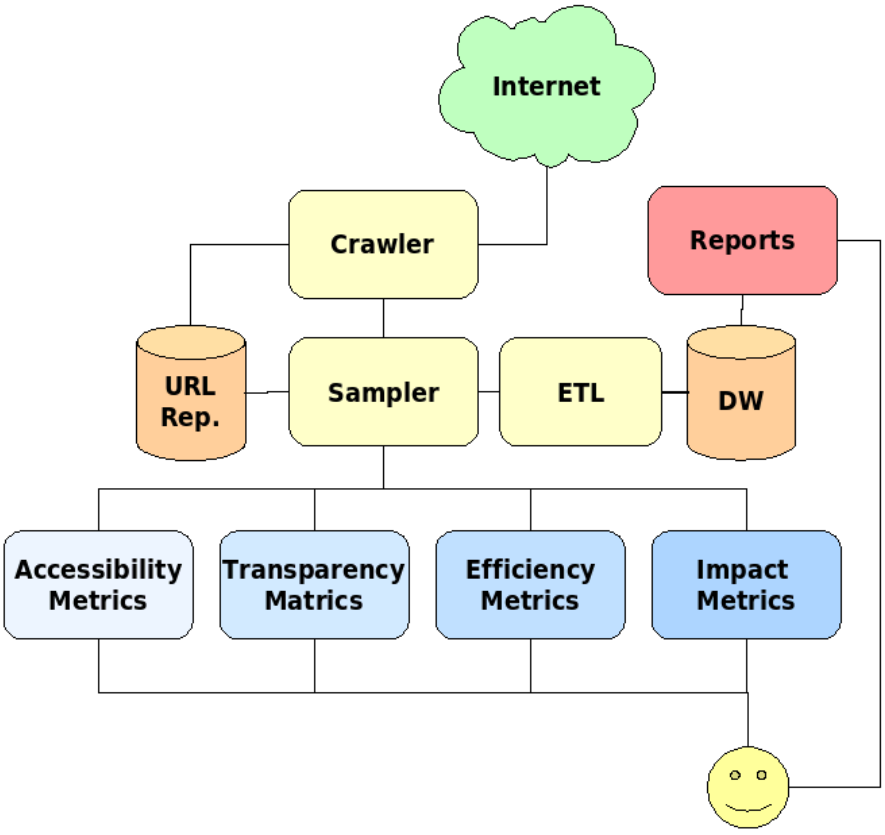

Figure 1: Proposed architecture for eGovMon

2. **Learning Algorithms**

   A key part of the Ph.D. is defining learning algorithms. Such an approach is expected to contain the following.

   - **Key Features.**

     Based on results from manual evaluation done by the municipalities involved in eGovMon key features will be extracted.

     For reason of clarity, we provide an example. Lets assume that part of the goal is to create an algorithm which automatically identifies an online post list on public web sites. There will be performed manual evaluations identifying such lists and providing links to these. In this case, a key feature of this may be (1) The actual words which are normally part of such a web page containing a post list, (2) number of PDF-documents available (post often contain PDF-documents listing

the interaction with the citizens) (3) number of dates available (4) oldest and newest date present and so on.

The essential part is to identify which key features are important and which are not.

- **Defining and implementing algorithm.**

  Based on the key features, an algorithm with the intent to e.g. classify a web page as post list or not post list will be created. The algorithms should to be created with both accuracy and speed in mind.

- **Validating the algorithm.**

  The accuracy of the developed algorithms will be validated. The already retrieved manual evaluation results can be used as training data. To verify the results, the algorithm could be run on the following two data sets:

  – New web pages.
    If this is approach is used, the results needs to be verified manually.
  – Already existing training data.
    In order to not give the algorithms an unfair advantage, the algorithms should never validated using the exact same data used for training.
    Such a validation is often done by ether dividing the available into separate two part, one used only for training and one for only verifying the algorithm. Often these subsets are selected at random and the validation is repeated a number of times.
    Another possibility, often used when limited amounts of training data is available, is a so-called leave-one-out validation[3]. In this case, the algorithm is trained with every available sample except one. The remaining sample will be used to validate the algorithm. This is repeated for every sample available.

  It should be noted that this process is expected to be iterative. It could be the case that during the validation phase, it is discovered that other key-features should be extracted.

3. **Large scale web evaluation.**

   The eGovMon project plans to do large scale web evaluation which in essence means that many web sites should be evaluated quickly. This provides some challenges to the entire architecture such as performance optimizations.

2.3.2 Scientific output and algorithms

The expected scientific output from this Ph.D. is divided into two main categories:

[3]Several variants of the leave-one-out validation exists, such as k-fold cross validation.

- An specification and implementation of an observatory which automatically benchmarks web sites according to accessibility, transparency, efficiency and impact.
- Empirical results of accessibility, transparency, efficiency and impact, both retrieved automatically using the observatory implementation and from manual surveys to public web site owners, such as municipalities.

The key algorithms in this survey are learning algorithms. Learning algorithms makes it possible to test for fairly advanced issues without the tedious process of defining the algorithms manually. The algorithms learns, based on some input data, what is the correct output with some precision.

Possibly the most known example of such an algorithm is e-mail junk classification. The learning algorithm automatically classifies incoming e-mail as ether junk or not junk. Furthermore, the criteria for an e-mail to be classified as junk is depending on the manual classifications done by the user. A consequence of this is that the classification accuracy is continuously increased as the number of e-mails increase.

Similarly, I could see the same potential in evaluation of eGovernment services. For reason of clarity, I present an example: It is good accessibility to provide alternative text to images that correctly describe the corresponding image[4]. It is however hard to test that the description is correct automatically and most evaluators are limited by only testing the presence of such alternative texts rather than validity of these. Automatic and semi-automatic web page generators (such as content management systems) sometimes add a default alternative text such as *image1* which clearly does not describe any corresponding image correctly and is thus inaccessible. Detecting all such words or combination of words is a tedious procedure. However, similar to junk mail classification described above we can clearly see the potential of automatically classify alternative texts as *image1* as inaccessible while alternative texts as *cat* as accessible.

Learning algorithms can in most cases not return results with a 100% accuracy. Since the algorithms often depend on training data which in most cases contains bias or is incomplete[5]. Despite of this, many such applied algorithms has a precision close to 100% such as a correctly trained e-mail junk classifier. Such filters will in almost every case correctly classify an e-mail as spam or not. However, occasionally, there will be e-mails which incorrectly goes through the filters.

Following I mention and shortly describe the learning algorithms methods which are likely to be applicable for this work.

Learning algorithms could be divided into top-level categories such as:

[4]As an example, in a web page about farm animals an image of a cat could have a corresponding alternative text: *cat* not *image1*.

[5]We could never expect the input data to be complete or perfect. If this was the case, there would be no need for learning algorithms as the perfect correct output could already be known by examining the data. To be put into perspective, in order to have a complete data junk mail classification, every exising junk mail needs to be known - which clearly is not viable

- **Supervised Learning:** Algorithms where some input is provided in order to approximate an output, such as an algorithm for claiming/classifying the presence of post lists.
- **Unsupervised Learning:** Algorithms where no direct output exists and the intention is rather to understand how the data is organized, such as clustering organizing of e-mail interaction.
- **Re-inforced learning:** Algorithms where each action influences the environment of which it is trying to learn, which again influences the algorithm, such as learning automata. In such an environment, each action taken is dependant on the previous actions taken.

2.3.3 Web-mining and Data Mining

- **Web Content Mining**: Analyzing the actual content of web pages, such as the actual text. This is likely to be the most applicable area of learning algorithms and could for example be used to analyze if whether the content of a web page resembles the content of a post list, which could further be used as input for a classification algorithm.
- **Web Structure Mining**: Analyzing of how the web is structured. In this field web sites are considered graphs, where each page is a node and the links between pages are one way connections between the nodes. This area could be valuable to e.g. analyze the navigability of a web site. This could further be an important input to measuring navigability (Guideline 13 - WCAG 1.0 [Wor99]).
- **Web Usage Mining**: Analyzing how a web site is used. This area typically analysis web site logs to see which pages within a web site is mostly used, and how the users actually using the web sites. If such logs are available, this could be a valuable input to measuring impact. An online service which has a large impact on the users could also be extracted in the web site log.

2.3.4 Classification

Classification is the area of claiming that some, not before seen, input data is part of an already known class, e.g. classifying a new e-mail as junk or not junk. A lot of research exists in the area and could be used in relation web web-content mining.

Several algorithm exists for this, including:

- **Nearest Neighbor**. For the nearest neighbor, the training data and input data are organized into points (typically a space with many dimensions). Any new input data is classified accordring to class of the closest neighbor (using e.g. euclidean distance) of the input data.

- **K-Nearest Neighbor**. In contrast to nearest neighbor, in this algorithm, the input data is classified based on a vote of the k nearest neighbors. K-nearest neighbor is equal to the nearest neighbor if $k = 1$.
- **Naïve Bayes** is a probabilistic classifier using bayes theorem for classification of the input data. For each available class it calculates the probability that the input data is part of this class (based selected features). The input data will be classified based on which class has the highest probability of being part of.
- **Maximum Likelyhood** utilized the known distribution of the training data (e.g. guassian distribution) and uses this to determine which of the training classes it is it most likely that the input data could have been drawn from.
- **Classification/Regression Tree** is typically used for data which is dependant (in contrast to e.g. Naïve Bayes which requires independent data). Here a tree is created based based on the features of the data. When classifying, the tree is traversed where each node is equivalent to a decision and feature. When a leaf node is reached, the decision is made.

2.4 Potential Significance

The significance for this project can be divided into two: technical significance and significance of the actual results.

2.4.1 Technical Significance

Survey in the areas part of this Ph.D. is traditionally done manually and there exists a significant potential by creating fully- or semi automatic algorithms for this. Some automatic tests used for benchmarking web content already exists, but these are mostly formal deterministic. Such tests could for example detect the presence of alternative texts for images, but cannot claim if the alternative text is correctly corresponding to the image. In other words, the existing tests are limited when it comes to what they can test.

By expending the automatable tests to also include heuristic tests, such as learning algorithms and pattern recognition algorithms, what is testable automatically will increase drastically.

The algorithms to be developed will be a significant contribution in the field of web-mining. Naturally, a lot of web-mining algorithms already exists. However, to the best of my knowledge, there exists very few algorithms applied in the area of eGovMon.

In the eGovMon project, manual surveys will be applied by the municipalities involved. The results from these can be used as training data for the algorithms and is therefor a valuable input to learning algorithms to be created.

2.4.2 Societal significance

All areas part of the eGovMon have received a lot of attention these last years. Presenting results retrieved by eGovMon could contribute to raising the awareness in all areas and contribute to more accessible web sites, more transparent and efficient public services and more impact for the citizens using these services.

Often the situation is *what gets measured gets done*. Following are two practical examples of this.

On a yearly basis DIFI (previously Norge.no) evaluates public Norwegian web sites [Dir08]. Their intention is to measure how accessibility, user friendly and informative each web site is. Each web sites receives a score from zero to six stars. The measurement results receives a lot of attention and it is a goal for many Norwegian municipalities to get many stars in this ranking.

Another well known survey is the Programme for International Student Assessment (PISA) [Org08]. Their measurements include student skills in reading, mathematics and science literacy. This survey has received so much attention in the media that it is even basis for political discussion in several countries.

Similarly, the eGovMon project with the contribution from this Ph.D., if it reaches its full potential, can reach similar significance on the society.

2.5 Time schedule

In this section I present the over all work schedule of the Ph.D. Note that several of these activities will be done in collaboration with other members of the eGovMon project.

2.5.1 2008

Following is the work scheduled for 2008, including 2008-12.

- Literature review of the indicators focusing on transparency.
- Write deliverable on requirement analysis of transparency.
- Complete courses scheduled for 2008 (see chapter 4).
- Hold presentation in NOKIOS conference [FN08].
- Participate at United Nations Expert Group Meeting on E-Government Readiness Indicators.
- Write two publications (see chapter 2.7).

2.5.2 2009 First Semester

Following is the work scheduled for 2009-01 to 2009-06.

- Complete courses scheduled for 2009 semester (see chapter 4).
- Attend planned conference (see chapter 6).

- Write one publication (see chapter 2.7).
- Write deliverable on requirement analysis of transparency.
- Write deliverable on definition of algorithms for transparency.
- Write specification and implementation of algorithms for transparency.
- Review results from transparency evaluations.
- Producing and presentation of results for transparency.
- Literature Review of the indicators focusing on Accessibility.

2.5.3 2009 Second Semester

Following is the work scheduled for 2009-07 to 2009-12.

- Complete courses scheduled for 2009 semester (see chapter 4).
- Attend planned conference (see chapter 6).
- Write one publication (see chapter 2.7).
- Write deliverable on requirement analysis for accessibility.
- Write deliverable on definition of algorithms for accessibility.
- Write specification and implementation of algorithms for accessibility.
- Review of results from transparency evaluations.
- Literature review of of the indicators focusing on efficiency.
- Write deliverable on requirement analysis of efficiency.
- Write deliverable on definition of algorithms of efficiency.
- Write specification and implementation of algorithms for efficiency.
- Review results from efficiency evaluations.
- Foreign stay for between three to six months - starting 2009-09.

2.5.4 2010 First Semester

Following is the work scheduled for 2010-01 to 2010-06.

- No courses planned for 2010 first semester.
- Attend planned conference (see chapter 6).
- Write one publication (see chapter 2.7).

- Foreign stay for between three to six months - ending 2010-03.
- Review results from efficiency evaluations.
- Producing and presentation of results for efficiency.
- Write deliverable on literature review of of the indicators focusing on impact.

2.5.5 2010 Second Semester

Following is the work scheduled for 2010-06 to 2010-10.

- No courses planned for 2010 second semester.
- Attend planned conference (see chapter 6).
- Write one publication (see chapter 2.7).
- Write deliverable on requirement analysis of impact.
- Write deliverable on definition of algorithms of impact.
- Write specification and implementation of algorithms for impact.
- Review results from impact evaluation

2.5.6 2011

Following is the work scheduled for 2011, including 2011-09.

- No courses planned for 2011.
- Attend planned conference (see chapter 6).
- Write one publication (see chapter 2.7).
- Producing and presentation of results for impact.
- Writing and finalizing of Thesis.

2.6 Outline of Content

The Ph.D. thesis will be a collection of papers produced during the Ph.D. project. The thesis will consist of an abstract,introduction,conclusion and one chapter for each of the papers presented in chapter 2.7.

2.7 Tentative/Titles on Papers

In this chapter I present potential publications part of the Ph.D. Note that the publications listed here are only potential publications. In this Ph.D., I plan to publish two publications yearly (one paper per semester). One of these publications is planned to be for a journal, the remaining will be for conferences.

Additionally, I plan to write and contribute to scientific reports for the eGovMon project.

The authors of the papers have deliberately been left out. Most papers will have several authors. However, issues like: who will author the different papers, what will the order of authorship be is not defined.

Naturally, it is a prerequisite that the Ph.D. student contributes significantly to all the papers part of the Ph.D. This means the student needs to be first author of most of the papers which is included in the final thesis.

2.7.1 Planned Publications for 2008

This section contains planned publications for 2008.

1. **Architecture for large-scale automatic web accessibility evaluation based on the UWEM methodology.** [UMOP+08]
   - **Status:** Submitted and accepted.
   - **Outlet:** Conference: Norsk Informatikkonferanse (NIK 2009).
2. **How Accessible is the European Web?**
   - **Status:** Planned.
   - **Outlet:** Popular Science: The Economist.

2.7.2 Planned Publications for 2009

This section contains planned publications for 2009.

1. **Comparing Manual Results from DIFI[Dir08], Expert Evaluation and automatic results from eGovMon**
   - **Status:** Planned.
   - **Collaboration with:** Universidad Politecnica de Madrid.
   - **Outlet:** Conference: 13th International Conference on Human Computer Interaction (HCI2009).
2. **Applying Learning Algorithms for benchmarking Transparency and Accessibility**
   - **Status:** Planned.
   - **Outlet:** Conference: 9th European Conference on eGovernment (EGEP 2009).

2.7.3 Planned Publications for 2010

This section contains planned publications for 2010.

1. **Evaluation of Public Web Sites Based on the Web Content Accessibility Guidelines 2.0**
   - **Status:** Planned.
   - **Outlet:** Conference: 13th International Conference on Computers Helping People with Special Needs (ICCHP 2010).
   - **Collaboration with:** Zürich University of Applied Sciences.
2. **Applying Learning Algorithms for Measuring Efficiency of Public Web Sites**
   - **Status:** Planned.
   - **Outlet:** Conference: eChallenges 2009.

2.7.4 Planned Publications for 2011

This section contains planned publications for 2011.

1. **Machine Learning Algorithms for Impact on the Web.**
   - **Status:** Planned.
   - **Outlet:** Conference: 20th International World Wide Web Conference.
2. **Accessibility, Transparency, Efficiency and Impact of the Public Web - Results**
   - **Status:** Planned.
   - **Outlet:** Journal: IEEE Journal on Internet Computing.

3 Agreement on relationships

For my Ph.D. I have two supervisors, main supervisor Mikael H. Snaprud and co-supervisor Christian S. Jensen.

3.1 Main supervisor

The main supervisor for this Ph.D. is Dr. Mikael H. Snaprud at Tingtun AS. The student and the supervisor will have a weekly physical meetings. Each meeting is scheduled to last for approximately one hour. It should be noted that the duration is clearly dependant on what needs to be discussed and the actual duration of each meeting will depend on this.
When physical meetings are not possible, such as during the stay at a foreign research institute, the communication will be done via e-mail and telephone.

3.2 Co-Supervisor

The co-supervisor for this Ph.D. is Prof. Christian S. Jensen at Aalborg University Department of Computer Science. The student and the supervisor will have regular interaction via e-mail. Additionally, there will be scheduled for both phone meetings and physical meetings when needed.

4 Ph.D. Courses

The following courses are planned to be followed. The periods are organized into the month the course will be finalized. Note that the decision to participate in some of the courses have not been finalized yet and the plan may be adjusted. Further note that some courses have already been concluded. Both planned and concluded courses are presented in this chapter.
In total the student should complete courses summing to 30 ECTS (equivalent to 6 months of work). From this 15 ECTS needs to be project related courses and an additional 15 ECTS needs to be joint study related courses. Of the project related courses, 5 ECTS can be achieved by attendance at conferences. For this to be accepted, the student needs to write a report or hold a presentation which needs to be accepted by the supervisors.

4.1 Project Related Courses and Conferences

Requirement: **15 ECTS**.

4.1.1 Project Related Courses

Course	ECTS	Period	Status
IKT620 - Ph.D. Course on Applied Pattern Recognition at University of Agder	2.0	2008-04	Pass[8]
Mining the Social Web - Web Mining for WEB2.0	2.0	2008-09	Pass[8]
Not yet defined	6.0		
Total	**10.0**		

Table 2: Overview of Project Related Courses

4.1.2 Conferences

Conference	ECTS	Period	Status
ICCHP 2008 Conference	1.0	2008-07	Running[6]
NOKIOS 2008 Conference	1.0	2007-10	Planned
NIK 2008 Conference	1.0	2008-11	Planned
HCI International 2009	1.0	2009-04	Planned
ICCHP 2010 Conference	1.0	2010-07	Planned
Total	**5.0**		

Table 3: Overview of Conferences that contribute to ECTS.

[6]The Conference has been attended. However, the required report has yet to be submitted.

4.2 Joint Study Courses

Requirement: **15 ECTS**.

Course	ECTS	Period	Status
Ph.D. Course in English for Research Fellows	2.0[7]	2007-11	Pass[8]
Writing and Reviewing Scientific Papers	3.75	2008-11	Ongoing
Theories of Science	2.5	2008-11	Planned
Design and Analysis of Experiments	4.0	2008-10	Ongoing
Yet to be defined	2.75		
Total	**15.0**		

Table 4: Overview of Joint Study Courses

[7]The number of ECTS for this course has not yet been defined.

[8]The course is only graded Pass or Fail where Pass is equivalent of the grade A or B, while Fail is equivalent to C,D,E or F.

5 Teaching Obligations

Teaching obligations are not applicable for this Ph.D.

6 Dissemination

6.1 Participation in Relevant Conferences

Participating at relevant conferences, including presentation the work, is a vital part for a Ph.D. student. These activities should add up to at least 3 months.

6.1.1 2008

Following are some of the conferences I have participated in and conferences I plan to contribute to in 2008.

- 11th International Conference on Computers Helping People with Special Needs, ICCHP 2008, including contribution of paper and presentation [BHN+08].
- Norsk Konferanse for IKT i offentlig sektor (NOKIOS), 15th-17th of October, 2008. `http://www.nokios.no/`.
- United Nations Public Administration Network Expert Group Meeting on e-Government Readiness Indicators: Getting to the Next Level. 11th-12th of December, 2008. Invitation received. [Uni08]

6.1.2 2009

Following are some of the conferences I plan to contribute to in 2009.

- 18th International World Wide Web Conference, April 20th-24th, 2009 `http://www2009.org/`. Submission deadline 3rd of November 2009.
- 13th Internation Conference of Human Computer Interaction (HCI), 19th-24th of July, 2009 `http://www.hci2009.org/`. Submission deadline 20th of October 2008.
- 9th European Conference on eGovernment (ECEG). 15th-16h June. `http://www.academic-conferences.org/eceg/eceg2009/eceg09-home.htm`.
- eChallenges 2009 `http://www.echallenges.org/`.

6.1.3 2010

Following are some of the conferences I plan to contribute to in 2010.

- 13th International Conference on Computers Helping People with Special Needs, ICCHP 2010.

- 19th International World Wide Web Conference.
- 10th European Conference on eGovernment (ECEG). 21st-22nd of June. http://www.academic-conferences.org/eceg/eceg-future.htm.
- eChallenges 2010 http://www.echallenges.org/.
- Journal of the American Society for Information Science and Technology http://www.asis.org/.

6.1.4 2011

Following are some of the conferences I plan to contribute to in 2011.

- IEEE Journal On Intelligent Systemshttp://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=9670.
- 20th International World Wide Web Conference.
- 14th Internation Conference of Human Computer Interaction (HCI).
- 10th European Conference on eGovernment (ECEG).
- eChallenges 2011 http://www.echallenges.org/.
- IEEE Journal On Internet Computinghttp://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=4236.

6.2 Other Dissemination Activities

- Technical Kickoff, University of Agder, September 11th-14th 2008.
- Workshop on making accessible web sites, Mandal, Norway, September 5 2008.
- Kick-off Meeting, Oslo, Norway, April 9-10 2008. In collaboration with the GoOpen conference[Go 08].

For more dissemination activities, see chapters describing external collaboration (8.2, 8.3 and 8.4).

7 Patents

As an external Ph.D. student, employed in the Norwegian company Tingtun AS, rights to all work (source code, publications etc.) are owned by Tingtun AS.

It should be noted that in the eGovMon project and this corresponding Ph.D., all scientific results will be released under an open license, and all software will be released as open source (GPL license or similar). This means all results from this Ph.D. will be in the public domain and there is no intention of filing for patents based on the work in this Ph.D.

8 External collaboration

8.1 Research Stay

At the end of the first year or beginning of second (approximately 2009-09) it is intended that I stay at a research institute abroad lasting for between three and six months. It is not yet decided which institute I will be connected to during this period. Possible locations include:

- Faculdad de Informatice,Universidad Politecnica de Madrid. `http://www.fi.upm.es/`.
- Trace Center. `http://trace.wisc.edu/`.
- Word Wide Web Consortium. `http://www.w3.org`.
- Bartimeus Accessibility Foundation, Holland `http://www.accessibility.nl/?languageId=2`.
- Informatics and Telematics Institute, Greece. `http://www.iti.gr/db.php/en/pages/about.html`.
- Foundation for Information Technology Accessibility (FITA), Malta. `http://www.knpd.org/mittsfita/`.
- Fundacion OVSI, Spain. `http://www.ovsi.com/`.
- Stakes, Finland. `http://www.stakes.fi/EN/index.htm`.
- IBM, Tokyo Research Laboratory. `http://www.research.ibm.com/trl/extfnt_e.htm`.
- United Nations, New York. `http://www.un.org/english/`.
- Organization for Economic Co-operation and Development (OECD). `http://www.oecd.org/`.
- World Bank. `http://www.worldbank.org/`.
- Indian Institute for Technology. `http://www.iitd.ac.in/`

8.2 eGovMon partners

The development team of the eGovMon project will have phone meetings three times a week collaborating tasks, challenges, potential improvements and other relevant development information. Additionally, we will have regular e-mail interactions. This collaboration will mainly happen between members of Tingtun AS, Forschungsinstitut Technologie und Behinderung (FTB) in Germany, University of Agder in Norway, Aalborg University in Denmark.

8.3 Municipalities

19 Norwegian municipalities are partners in the eGovMon project. There will exist regular collaborations between the municipalities and the technical eGovMon team, such as phone meetings, workshops etc.

8.4 United Nations

As part of the Ph.D. I have been invited to the United Nations Expert Group Meeting on E-Government Readiness Indicators: Getting to the Next Level [Uni08]. There will be a physcal meeting in 2008-12. Further collaboration after this point is expected.

9 Finance Budget

The Ph.D. is financed through the eGovMon [Sna08a] project which is co-funded by the research council of Norway. I will be an external Ph.D. at Aalborg University. In this period I will be employed at the Norwegian company Tingtun AS [Sna].

The yearly payment from Tingtun AS to Aalborg University is outlined in table 5 in DKK including 20% overhead.

Activity	**Cost**	
Supervision	30 660	DKK
Courses	11 040	DKK
Foreign stay	0	DKK
VIP administration	6 360	DKK
Evaluation	12 960	DKK
Sum	**61 020**	**DKK**

Table 5: Payment to Aalborg University

References

[Ber08] Lasse Berntzen. e-transparency of political decision making, 2008. [Unpublished].

[BHN+08] Christian Bühler, Helmut Heck, Annika Nietzio, Morten Goodwin Olsen, and Mikael Snaprud. Monitoring accessibility of governmental web sites in europe. In *International Conference on Computers Helping People with Special Needs.*, pages 410–417, 2008.

[CKM07] Kevin Cullen, Lutz Kubitschke, and Ingo Meyer. Assessment of the status of eAccessibility in Europe. Available online at: `http://ec.europa.eu/information_society/activities/einclusion/library/studies/meac_study/index_en.htm`, 2007.

[Dir08] Direktorat for Forvaltning og IKT. Kvalitetsvurdering av offentlige nettsteder. Available online at: `http://www.norge.no/kvalitet/`, October 2008. [Online; accessed 9-October-2008].

[Eur05] European Commission. i2010 - A European information society for growth and employment. Available online at: `http://europa.eu.int/information_society/eeurope/i2010/docs/communications/com_229_i2010_310505_fv_en.pdf`, 2005.

[FN08] Fornyings- og Administrasjonsdepartementet and Norges Forskningsrd. Norsk konferanse for ikt i offentlig sektor (nokios). Available online at: `http://www.nokios.no/start`, October 2008. [Online; accessed 9-October-2008].

[Go 08] Go Open. Go open 2008. Available online at: `http://www.friprog.no/ez/`, October 2008. [Online; accessed 9-October-2008].

[Hee06] Richard Heeks. Understanding and measuring egovernment: International benchmarking studies. Available online at: `http://unpan1.un.org/intradoc/groups/public/documents/UN/UNPAN023686.pdf`, 2006.

[Mil08] Jeremy Millard. egovernment measurement for policy makers, August 2008.

[Nom06] Nomensa. United Nations global audit of web accessibility. Available online at: `http://www.un.org/esa/socdev/enable/documents/fnomensarep.pdf`, 2006.

[NSV08] Annika Nietzio, Christophe Strobbe, and Erik Velleman. The unified web evaluation methodology (uwem) 1.2 for wcag 1.0. In *International Conference on Computers Helping People with Special Needs.*, pages 394–401, 2008.

[NUMGO07] Annika Nietzio, Nils Ulltveit-Moe, Terje Gjœsæter, and Morten Goodwin Olsen. Unified Web Evaluation Methodology Indicator Refinement. Available online at: `http://www.wabcluster.org/uwem1_1/`, 2007.

[Org08] Organisation for Economic Co-operation and Development. Programme for international student assessment. Available online at: `http://www.pisa.oecd.org/pages/0,3417,en_32252351_32235731_1_1_1_1_1,00.html`, October 2008.

[Osi08] David Osimo. Benchmarking egovernment in the web 2.0 era: what to measure, and how. Available online at: `http://www.epracticejournal.eu/document/4972`, August 2008.

[OUMN+08] Morten Goodwin Olsen, Nils Ulltveit-Moe, Annika Nietzio, Mikal Snaprud, and Anand B. Pillai. Test report for eiao version 2.2 of the observatory software. Available online at: `http://eiao.net/resources`, 2008. [Online; accessed 9-October-2008].

[Ove08] Overheid heeft Antwoord. The web guidelines quality model. Available online at: `http://www.webrichtlijnen.nl/english/`, October 2008.

[Pav06] Jan Pavel. Public procurement market transparency indicators in the czech republic. Available online at: `http://www.transparency.cz/pdf/vz_index_transparentnosti_metodika_en.pdf`, October 2006.

[Sna] Mikael Snaprud. Tingtun. Available online at: `http://www.tingtun.no`. [Online; accessed 10-October-2008].

[Sna08a] Mikael Snaprud. eGovernment Monitoring. Available online at: `http://www.egovmon.no`, October 2008. [Online; accessed 9-October-2008].

[Sna08b] Mikael Snaprud. European Internet Accessibiliby Observatory. Available online at: `http://www.eiao.net`, October 2008. [Online; accessed 9-October-2008].

[SOA06] Mikael Snaprud, Morten Goodwin Olsen, and Finn Aslaksen. Automatic benchmarking and presentation of the first results from the european internet accessibility observatory. Available online at: `http://www.mortengoodwin.net/publicationfiles/echallenges2006.pdf/document_view`, October 2006.

[The03a] The UK Voluntary Sector Research Group. Measuring impact - a guide to resources. Available online at: `http://www.ncvo-vol.org.uk/asp/uploads/uploadedfiles/1/455/measuringimpactguide.pdf`, 2003.

[The03b] The UK Voluntary Sector Research Group. Measuring impact - case-sudies of impact assessment in small and medium-size voluntary organisations. Available online at: `http://www.ncvo-vol.org.uk/asp/uploads/uploadedfiles/1/456/measuringimpactreport.pdf`, 2003.

[Tra08] Transparency International. Global corruption report - corruption in the water sector. Available online at: `http://www.transparency.org/publications/gcr/download_gcr`, 2008.

[UK 05] UK Cabinet Office. eAccessibility of public sector services in the European Union. Available online at: `http://www.cabinetoffice.gov.uk/e-government/eaccessibility`, 2005. [Online; accessed 9-October-2008].

[UMOP+08] Nils Ulltveit-Moe, Morten Goodwin Olsen, Anand B. Pillai, Christian Thomsen, Terje Gjster, and Mikael Snaprud. Architecture for large-scale automatic web accessibility evaluation based on the uwem methodology, 2008.

[Uni08] United Nations Public Administration Network. Expert group meeting - e-government readiness indicators: Getting to the next level. Available online at: `http://www.unpan.org/Regions/Global/Events/Conferences/tabid/458/mctl/EventDetails/ModuleID/1510/ItemID/1010/Default.aspx`, December 2008.

[Web07] Web Accessibility Benchmarking Cluster. Unified Web Evaluation Methodology (uwem 1.2). Available online at: `http://www.wabcluster.org/uwem1_2/`, 2007.

[Wik08] Wikipedia. Web 2.0 — wikipedia, the free encyclopedia, 2008. [Online; accessed 9-October-2008].

[Wor99] World Wide Web Consortium. Web Content Accessibility Guidelines 1.0. W3C recommendation 5 may 1999. Available online at: `http://www.w3.org/TR/WCAG10/`, 1999.

[Wor08] World Wide Web Consortium. Web Content Accessibility Guidelines 2.0. W3C release candidata. Available online at: `http://www.w3.org/TR/WCAG20/`, 2008.