

# Preface

A cornerstone of Web project management is sound effort estimation, which is the process by which a person, group of people, tool, model, or a combination of these predicts the amount of effort needed to accomplish a given task. The importance of having realistic effort estimates at an early stage in a project's life cycle is widely recognised by Web project managers and developers since accurate estimates are paramount to manage resources effectively, and to considerably increase the probability that projects will be finished on time and within budget.

However, recent findings showed that numerous Web projects worldwide are still not finished within time and budget, and one of the main reasons for that is the use of poor project management practices, which includes poor effort estimation (Ginige, 2002).

Given that the field of Web effort estimation is relatively new, with its first paper published only in 2000 (Mendes, Counsell, & Mosley, 2000), it would not be surprising that research findings in this field may not have reached industry widely.

Effort estimation, also known as cost estimation, is a necessary part of an effective process, whether this process is Web authoring, design, testing, or development as a whole. An effort estimation process involves the following steps (Fenton & Pfleeger, 1987).

1. The identification of factors perceived to influence effort (e.g., size measures, cost drivers) and the type of relationship they have with effort
2. The gathering of past project data using as basis the factors identified in Step 1. If gathering a Web company's own data is not possible, this phase can in-

volve the use of large cross-company data sets with data on numerous diverse projects.

3. The use of one or several effort estimation techniques to be applied to estimate effort for new projects. The effort estimates generated by these techniques can also be adjusted using expert opinion.
4. The assessment of how effective the prediction technique is

It is important that all these steps are followed; otherwise, the effort estimates obtained may not be as useful as they should.

Another important point is that effort prediction is generally based on the following.

- Knowledge of previous similar projects and applications managed by an organisation. This means that data from past projects may be essential in helping organisations estimate effort for new projects.
- Other project and application characteristics that are believed to be related to effort. Examples of project characteristics are the size of the development team, its experience with working on similar projects, and the number of different programming languages being used in a project. An example of an application characteristic is the size of the problem to be developed, which may encompass different attributes such as the total number of new Web pages, total number of images, number of features that were reused, and so forth.

The objective of this book is therefore to provide Web companies, researchers, and students with the necessary knowledge on Web effort and cost estimation. It includes step-by-step guidelines on how to use and compare several effort estimation techniques, which may considerably help companies improve their current effort estimation practices, and help researchers and students understand the process that needs to be carried out to estimate development effort.

The effort estimation techniques that are detailed in this book are those that have been to date the three mostly used effort estimation techniques in the Web effort estimation literature, namely, stepwise regression, case-based reasoning, and classification and regression trees (CART).

Throughout this book, we take the view that the use of past data on finished Web projects can be extremely useful and necessary to help obtain accurate effort estimates for new Web projects and also to help Web companies understand how they currently estimate effort for their new projects.

All chapters are self-contained, and whenever applicable, the chapters present a literature review of previous studies on the topic being explained in order to ensure that readers are familiarised with what has been previously published in the literature.

We also present detailed chapters on principles of statistics and empirical studies to provide readers with additional knowledge not only useful for Web effort and cost estimation, but also useful within the context of Web engineering.

It is important to note that despite Web effort estimation being one of the main components of project management, there are at least two other components also related to effort estimation that can indirectly benefit from the use of more sound estimating processes. These two other components are project productivity and development processes. Once a Web company identifies important factors that have a bearing on effort and gathers data on past projects, it is also possible to identify current productivity trends between projects and also between developers, and to assess if differences in development processes have any effect on the effort estimates proposed and learn how to improve processes. Therefore, a change to a more systematic way of estimating effort can also drive a company to improve its productivity and current development processes, which is always a positive outcome.

## Purpose

---

The purpose of this book is to introduce practitioners, lecturers, researchers, and students to Web effort estimation concepts and detailed case studies. Our objective is to provide detailed knowledge on Web effort estimation, step-by-step guidelines on how to use particular effort estimation techniques to estimate effort for new projects, and lessons on how to compare the effort estimates provided by these techniques, which may also include comparison with other benchmarking effort estimates.

The motivation for this book was threefold. First, our experience in dealing with Web companies to improve their effort estimates showed that companies would like to know how to use different effort estimation techniques and how to compare these techniques; however, the literature available was either too brief or unavailable. Second, our experience teaching Web effort estimation to postgraduate students showed that the books that were completely devoted to the topic of effort estimation were all applied to conventional software development projects, which are very different from Web projects (see Chapter 1 for a detailed discussion). Third, our experience in giving tutorials on Web effort estimation to Web engineering researchers also showed the need for a single place where all the necessary information was available. There are several papers on Web effort estimation available, and also a book chapter; however, none details the processes that are used by different techniques to obtain effort estimates. Papers may compare different techniques, however each technique is presented very briefly and readers have to look for complementary literature to understand in detail and learn how to use these techniques.

---

## Target Audience

---

The target audience of this book comprises Web project managers and developers, Web engineering and software engineering students, and Web and software researchers. This book does not assume readers are familiar with either effort estimation concepts or any of the effort estimation techniques described, or statistical principles.

---

## Outline

---

The book is organised into 11 chapters, each briefly described as follows.

Chapter I introduces terminology related to hypertext, since this is the model the Web is based upon, then provides an overview of differences between Web and software development with respect to their development processes, technologies, quality factors, and measures. Finally it discusses the differences between Web cost estimation and software cost estimation.

Chapter II introduces the concepts related to Web effort estimation and effort estimation techniques. These concepts are later used in further chapters.

Chapter III describes the process to be used to assess the accuracy of an effort technique or model. This process is called cross-validation. In parallel with conducting a cross-validation, prediction accuracy measures are also obtained and aggregated. Examples of accuracy measures are the mean magnitude of relative error (MMRE), the median magnitude of relative error (MdMRE), the magnitude of relative error (MRE), and the prediction at Level  $l$  (Pred[25]).

Chapter IV presents a literature survey of size measures (attributes) that have been published within the last 14 years and classifies the surveyed measures according to a proposed taxonomy. In addition, this chapter also discusses ways in which Web companies can devise their own size measures.

Chapter V presents a case study where a real effort prediction model based on data from completed industrial Web projects is constructed step by step using a statistical technique called regression analysis.

Chapter VI presents a case study that details step by step how to obtain effort estimations using real data from completed industrial Web projects using a machine learning technique called case-based reasoning.

Chapter VII presents a case study where a real effort prediction model based on data from completed industrial Web projects is constructed step by step using a machine learning technique called CART.

Chapter VIII details the use of statistical significance tests to compare different effort estimation techniques and models.

Chapter IX provides suggestions believed to help improve effort estimation practices that can be of benefit to Web companies, in particular, small Web development companies. The discussion also includes suggestions on how to improve project management practices as means to improving effort estimates.

Chapter X provides an introduction to parts of statistics that are frequently used when dealing with data for effort estimation. The concepts presented here are in no way exhaustive since statistics comprises a very large body of knowledge where entire books are devoted to specific topics. The parts that are the focus of this chapter are those that are necessary to use when building effort estimation models, and also when comparing different effort estimation techniques.

Chapter XI discusses the need for empirical investigations in Web engineering, which is motivated by the very definition of engineering. It describes the three main types of empirical investigations: surveys, case studies, and formal experiments. Although all three types of empirical investigations are presented, formal experiments are detailed further since they are the most difficult type of investigation to plan.

## References

---

- Fenton, N. E., & Pfleeger, S. L. (1997). *Software metrics: A rigorous and practical approach* (2<sup>nd</sup> ed.). Boston: PWS Publishing Company.
- Ginige, A. (2002, July). Workshop on Web engineering. Web engineering: Managing the complexity of Web systems development. *Proceedings of the 14<sup>th</sup> International Conference on Software Engineering and Knowledge Engineering* (pp. 72-729).
- Mendes, E., Counsell, S., & Mosley, N. (2000, June). Measurement and effort prediction of Web applications. *Proceedings of 2<sup>nd</sup> ICSE Workshop on Web Engineering* (pp. 57-74).