Chapter II
Web Development Effort Estimation: An Empirical Analysis

Emilia Mendes
University of Auckland, New Zealand

Silvia Abrahão
Valencia University of Technology, Spain

ABSTRACT

Effort models and effort estimates help project managers allocate resources, control costs and schedule, and improve current practices, leading to projects that are finished on time and within budget. In the context of Web development and maintenance, these issues are also crucial, and very challenging, given that Web projects have short schedules and a highly fluidic scope. Therefore, the objective of this chapter is to introduce the concepts related to Web effort estimation and effort estimation techniques. In addition, this chapter also details and compares, by means of a case study, three effort estimation techniques, chosen for this chapter because they have been to date the ones mostly used for Web effort estimation: Multivariate regression, Case-based reasoning, and Classification and Regression Trees. The case study uses data on industrial Web projects from Spanish Web companies.

INTRODUCTION

The Web is used as a delivery platform for numerous types of Web applications, ranging from complex e-commerce solutions with back-end databases to online personal static Web pages (Mendes, Mosley, & Counsell, 2005a). With the sheer diversity of Web application types and technologies employed, there are a growing number of Web companies bidding for as many Web projects as they can accommodate. As usual, in order to win the bid, companies estimate unrealistic
Web Development Effort Estimation

schedules, leading to applications that are rarely developed within time and budget (Mendes & Mosley, 2005).

The purpose of estimating effort is to predict the necessary amount of labour units to accomplish a given task, based on knowledge of previous similar projects and other project characteristics that are believed to be related to effort. Project characteristics (independent variables) are the input, and effort (dependent variable) is the output we wish to predict (see Figure 1). For example, a given Web company may find that to predict the effort necessary to implement a new Web application, it will require the following input: estimated number of new Web pages, total number of developers who will help develop the new Web application, developers’ average number of years of experience with the development tools employed, main programming language used, the number of functions/features (e.g., shopping cart) to be offered by the new Web application. Of these variables, estimated number of new Web pages and the number of functions/features to be offered by the new Web application are size variables (size measures); the other three, total number of developers who will help develop the new Web application, developers’ average number of years of experience with the development tools employed, and main programming language used, are not used to “size” the “problem to be solved” (Web application), but they are believed to influence the amount of effort necessary to develop a Web application, and in this sense are related to effort. Therefore, they are also considered input, and jointly named “cost drivers.”

The challenge in estimating effort is to obtain an estimate that is similar to the real amount of effort necessary to develop an application. Thus, research in this field aims to quantify and to determine the factors necessary to derive an estimate, such that the process of estimating effort can be fully understood, and can be repeated. In addition, it also uses and compares effort estimation techniques, looking for the technique(s) that provides most accurate effort estimates.

A task to be estimated can be as simple as developing a single function (e.g., creating a Web form with 10 fields) or as complex as developing a large application, and in general the one input (independent variable) assumed to have the strongest influence on effort is size. Cost drivers are also influential.

It is also often the case where knowledge about past projects, or even data on past finished projects, is also used to help derive an effort estimate (see Figure 1). The effort estimation techniques described later in this chapter all use data on past projects; however, many Web companies use as basis for their estimations solely knowledge on past projects, estimated size, and cost drivers.

Cost and effort are often used interchangeably within the context of effort estimation (prediction) because effort is taken as the main component of project costs. However, given that project costs also take into account other factors such as contingency and profit (Kitchenham, Pickard, Linkman, & Jones, 2003) we will use the word “effort” and not “cost” throughout this chapter.

The remainder of this chapter is structured as follows: Section 2 presents an introduction to Web effort estimation techniques, and a literature review and comparison of previous work in Web effort estimation. Section 3 presents a case study where a data set of industrial Web projects is used with three different effort estimation techniques to obtain effort estimates for these projects. This section also compares the prediction accuracy between the techniques. Section 4 discusses the necessary steps that need to be carried out.