Nowadays many web sites consist of hundreds, thousands or even hundreds of thousands of pages. The information on these pages can be used to answer many different questions, but at the same time the large number of pages makes it difficult for users to locate relevant information within a site. To assist users in their search, web sites offer a range of navigation means, such as in-text links and site search engines. In this thesis we focus on one of the oldest and most frequently used navigation means: hierarchical menus. In particular, we investigate how hierarchical menus can automatically be optimized in such a way that navigation becomes as efficient as possible.

Menus can support various aspects of navigation. Most menus are aimed at users with specific information needs. The goal of this type of menus is to help users to reach the information they need as fast as possible. Other menus are meant for users who do not know exactly what information they are looking for. These problem-oriented menus guide users step by step through the site. They first show pages with introductory information the users need. Once the users have determined what information they need, the menu guides them to more specific pages.

Menus consist of two elements: the structure of the menu hierarchy and the descriptions of the menu items. Both elements influence the efficiency of the users’ navigation. The structure of the hierarchy determines how many navigation steps are needed to reach certain pages. Menus that are not well-structured (for instance, because popular targets are located deep in the hierarchy) require users to make many navigation steps. Users read the descriptions to choose which menu items they will open. Accurate descriptions enable users to choose items that lead to the information they need. However, when the descriptions are unclear, users cannot make the right choices and often open incorrect items.

Until now menu optimization research was restricted to users with specific information needs and mainly covered the structures of menus. This thesis extends this work in several ways. We show that existing methods for menu optimization under certain circumstances do not result in the most efficient menu structures. We propose a method to overcome this problem. In addition, we address novel tasks, such as automatically creating problem-oriented menus and optimizing descriptions of links.

Various information sources can be exploited to determine how a menu needs to be
adapted. The most frequently used sources are the contents of the pages (the words), annotations that are manually added to the pages, and log files that contain data about the usage of the site. The methods that are presented in this thesis apply only usage data. Usage data have a number of advantages compared to data from other sources. Firstly, no human effort is needed to collect these data. Secondly, usage data can be used also on sites that consists largely of images or movies, while content-based methods are restricted to text. Finally, most methods that make use of content or annotations depend on domain-specific characteristics of sites. Usage-based methods are domain independent.

The first chapters of this thesis directly build on existing research on menu optimization. In these chapters we address the optimization of the structure of menus for users with specific information needs. We discover that existing methods that aim to minimize the average number of navigation steps, in fact do not accomplish this. Instead, in every navigation step, they maximize the probability that in that step the user will reach the information he needs. We show the effects of this discrepancy in theory and practice. Another weakness of these methods is that they make assumptions about the users of the site without verifying whether these assumptions hold. Our experiments demonstrate that this is a serious shortcoming as incorrect assumptions often lead to inefficient menus.

We present a framework that shows the assumptions underlying menu optimization methods in an organized way. The framework clearly shows that there are large differences between the assumptions that are made by the various menu optimization methods. For example, some methods assume that users read all menu items before making a choice, while others assume that users open an item as soon as an acceptable item is encountered. In total the framework distinguishes fifteen types of assumptions. We provide a method to systematically test the validity of the assumptions in the framework in the context of a particular site and its menu. For every possible combination of assumptions the method creates a model that can be used to predict user navigation through the menu. The predictions of the models are compared to the user navigation that is recorded in the log files. In this way, for every site we can find the model that is most consistent with its user population.

A model of a user population can be used to optimize a menu. We demonstrate a method to gradually improve a menu during a number of optimization steps. In each step the method applies various adaptation operations to the current version of the menu. This results in a number of alternative menu structures. The model is used to predict how much time the average user will need to reach the information he is looking for when using each of the alternative menus. The best structure is saved and used as starting point for the next optimization step. We evaluate the method in four case studies. The results are positive: the adaptations reduce the predicted navigation time, while the coherence of the menus is maintained.

Problem-oriented menus present users the information on a site in the right order. Therefore, to automatically construct these menus we need to know not only which pages users want to visit, but also in which order they want to read the pages. We present a model to describe the preferred reading order of a set of pages. A method is developed to the fill in the data of the model for a specific site on the basis of usage
Summary

data. Experiments with two sites show that this method can accurately determine the order in which users tend to read the pages of the sites. Moreover, on the basis of these models we can construct effective problem-oriented menus.

The last part of this thesis focuses on the optimization of descriptions of menu items. We present a model that describes how users navigate in the presence of accurate link descriptions as well as various types of inaccurate descriptions. Comparing the model to the users' navigation at certain locations in a menu allows us to determine the quality of the descriptions in the menu. We distinguish various types of inaccurate descriptions and for each type we provide a number of possible solutions. In an evaluation study this method was applied to the menus of three web sites. Webmasters of the sites judged that the findings of the method were very useful for improving the menus. On top of that, the adaptations to the menus that the webmasters chose on the basis of these findings significantly reduced the number of navigation mistakes.

In this thesis we showed how we can construct models on the basis of usage data and how these models can be used to improve the efficiency of hierarchical menus. Our model-based approach proved very effective for various aspects of menu optimization and in various domains. However, the benefits of models are not inherently limited to hierarchical menus or even to web sites. We believe, therefore, that in the future the model-based approach will also prove useful for the optimization of other types of interfaces.