Abstract. In our chapter we are working within the field of Web content mining. In relation to the user's description of a Web page, we define a new term: Named object. Named objects are used for a new classification of selected methods dealing with mining information from Web pages. This classification has been made on the basis of a survey of published methods. Our approach is based on the perception of a Web page through an intention. This intention is important both for the users and authors of a Web page. Named object is near to Web design patterns, which became a basis for our own mining method, Pattrio. The Pattrio method is introduced in this work together with a few experiments.

1 Introduction

A Web page is like a family house. Each of its parts has its purpose, determined by a function which it serves. Every part can be named so that all users envision approximately the same thing under that name such as living room, bathroom, lobby, bedroom, kitchen and balcony. In order for the inhabitants to orientate well in the house, certain rules are kept. From the point of view of these rules, all houses are similar. That is why it is usually not a problem for first time visitors to orientate in the house. We can describe the house quite precisely thanks to names. If we add information about a more detailed location such as sizes, colors, furnishings and further details to the description, then the future visitor can get an almost perfect notion of what he will see in the house when he comes in for the first time. We can also take an approach similar to description a building other than a family house (school, supermarket, office etc.) Also in this case the same applies for visitors and it is usually not a problem to orientate (of course it does not always have to be the case, as there are bad Web pages, there are also bad buildings).

Let us look at the problem from the other perspective. If we visit a building with a blindfolded person, then we can submit three basic tasks. The first is to find out what the purpose of the building is. The second is to find out what parts (e.g. rooms) the building contains. And the third task can be linked to the furnishings of individual rooms. When solving these tasks, it is probably possible to start with any of them. There is one more important issue. If the visitor completes some of
the tasks and we will require him to describe the result, he will certainly use
commonly used names, which describe the type of building, its parts and finally,
its furnishings.

Architect Christopher Alexander [1] brought in a similar and to a certain extent
formalized way of description. In our chapter in this book, we try to work with a
Web page in a similar way. And we try to show that this way of looking at a Web
page can moreover, be a good tool for the classification of some approaches in the
field of Web content mining. We can verify it is reasonable to use individually
named (labeled) part of a Web page to describe entire page. This holds true both
for the suggestion of methods for page semantics detection and for the technically
utilizable user's page description.

Our chapter is organized in the following way. In the second section, basic
principles concerning Web usability are described. In the third section, we will
explain what is meant by Web content mining and what typical tasks are dealt
with in this area. In the fourth section of the chapter, we will explain in detail what
is meant by the term intention in relation to Web page content. In this section, we
will also introduce a new term: Named object, as a basic abstraction related to the
intention. The fifth section of the chapter is devoted to the survey of approaches
which in some way relate to our view on a Web page. In the sixth section, we
present our method Pattrio, which is focused on the detection of Named objects.
We will describe experiments related to the successfulness of this method's
usability and to its results for partial tasks. The last sections of the chapter are
devoted to experiments, a summary and prospects for further research.

2 Web Usability

Web usability is closely linked to User Centered Design (UCD). In a wider sense,
UCD is a philosophy which results in the process of software system
development. The main difference from other approaches is that UCD tries to
optimize the user interface, so that it

- Corresponds to what users are used to
- Does not make the user change their way of working

Jacob Nielsen defines usability as follows: “Usability is a quality attribute that
assesses how easy user interfaces are to use.” In the book [29] many tests with
users are shown and many important recommendations for Web page creators
come out of the results of these tests.

However, one of the problems is that recommendations in the field of Web
usability are not completely formalized. In [14] a conclusion is formulated that
recommendations resulting from Web usability can be formalized with the use of
patterns (for more about patterns see section Web Design Patterns). One of the
pattern characteristics which describe verified experience is that they have an apt
name that characterizes the solved task. The methodology of using patterns for
Web design is very thoroughly elaborated in [8]. There we can find classification
of patterns and important recommendations for developers.