

Web Usage Mining

Abstract

The internet is made up of three different types of data: content, structure data and usage data. While most people originally lean towards seeing the content as the main information that is provided on the web, there is another side, the usage data, that explains how we use the web. Web usage mining looks to investigate the patterns in usage that help us understand what the users are interested in and what the internet is truly used for. By analyzing the way we use the web we can then begin to mold the internet and personalize it so that each user can have their own experience based on their own interests. A lot of this is done through web usage mining, which consists of taking the raw server logs and then pruning out the relevant data and then doing certain website analytics, association rule generation, sequential pattern analysis, and clustering and classification techniques. These things begin to give us a numerical way to analyze the relationships between the users, products and web pages that exist on the internet.

Web usage mining, while a very interesting topic, has somewhat fallen by the wayside in recent years to more narrowly focused segments data mining of web usage data, such as recommender systems for suggesting products, movies, friends, and more, social network mining for finding either specific information about a person or on the flip side looking at the general trends in the population of these online networking sites like Facebook, MySpace and others, and many other forms. Web usage mining, while not popular by name today, is still very relevant and in fact is driving the evolution of the internet as we use it.