

Abstract

Personalization is a key feature of the Web, which has facilitated rich user - and context - specific services. Many companies use personalization to improve their services, such as, search engine providers and web advertising networks.

Differentiating the information presented to users based on context and history is often desirable, However in most cases the users are unable to have control over how this differentiation is enabled and how it is manifested. The main concern is that personalization is not limited to search engines results and advertisements. Recent research results have detected evidence that the tracking infrastructure employed for personalization and targeted advertizing is also used to deploy price discrimination on various e-commerce web-sites.

The first step towards personalization is to be able to compile a detailed user profile. Profiling users is a complex task that requires collaboration within a network of third party domains and free services that exchange information based on digital traces.

For researchers, these profiling networks are black boxes, because they don't have access to their internal algorithms and the type of personal information they collect. The only way to infer how this networks operate is to control their input and then observe the output that they produce (filtered search results, products price difference, targeted ads, etc.).

To get a better insight on how this tracking infrastructure monitors Internet users and builds their profiles we propose a large scale distributed system that attempts to reveal common data sharing practices among different domains. The system helps users to detect price discrimination on a large number of online stores. The system is designed based on an open architecture so that it can easily evolve over time offering more privacy control to the end-user. In addition, the system encourages users to contribute information to the system, thus setting the foundations to create a large community around it.

In this thesis we describe the design and implementation of this system, presenting its objectives, architecture, key components, and future extensions.