Web Analytics

Information Retrieval and Web Analytics

upf. Universitat Pompeu Fabra Barcelona

Our Goal

Apply web analytics to our search engine web application.

1) Review Web Technology Basic
   - Web Servers and Browsers
   - Data Persistence for Analytics
2) Review Web Analytics
   - Web Analytics Features
   - Tracking Website Usage
   - Displaying Usage Statistics
3) Implement a Web Application
   - Python Web Framework
   - Simple Web App Structure
   - Data Model
   - Search Engine Integration

Web Technology Basics

upf. Universitat Pompeu Fabra Barcelona

Web Technology Basics

- HTTP protocol
  - client/server
  - request/response
- Web Server
  - HTTP server
  - Server side (backend) logic: Java, DotNet, Python, Ruby, NodeJS
  - Connects to persistence store (database, file system, etc)
  - Expose REST APIs
- Browser
  - HTTP client
  - Client side logic
  - JavaScript Engine
  - Document Object Model (DOM)
- JavaScript Technology
  - ECMAScript - standardization of JavaScript
  - TypeScript - typed JS, objects, classes
  - Angular, React, Vue

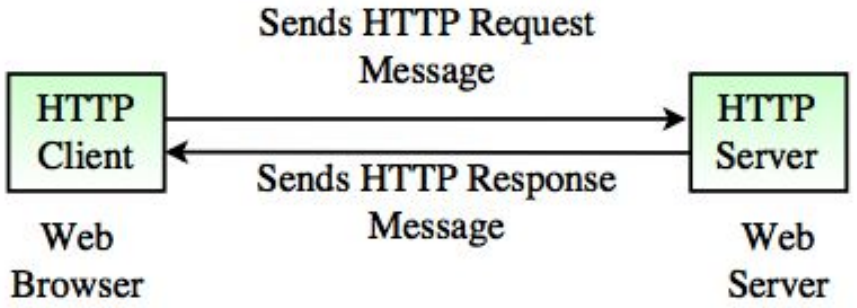


Fig. HTTP Protocol

Web Technology Basics

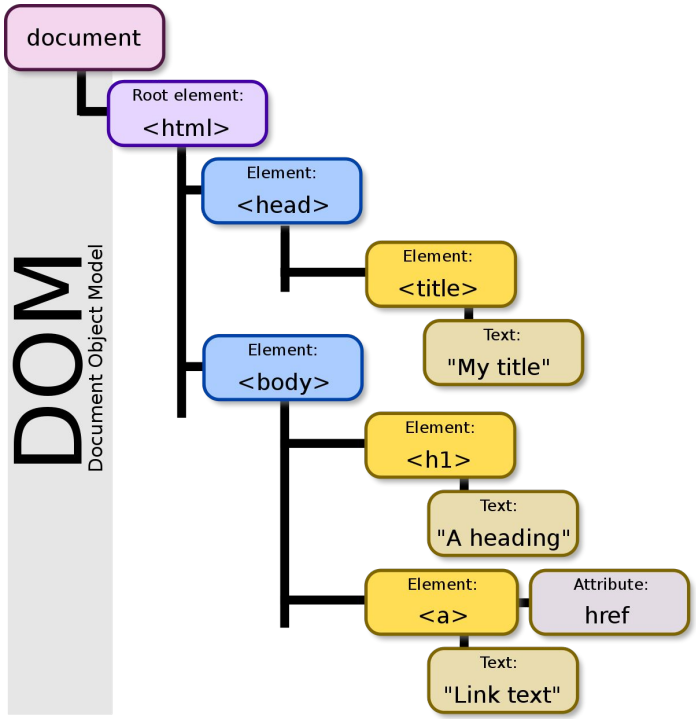
When a web page is loaded, **the browser creates** a **Document Object Model** of the page.

The **HTML DOM** is an Object Model for HTML:

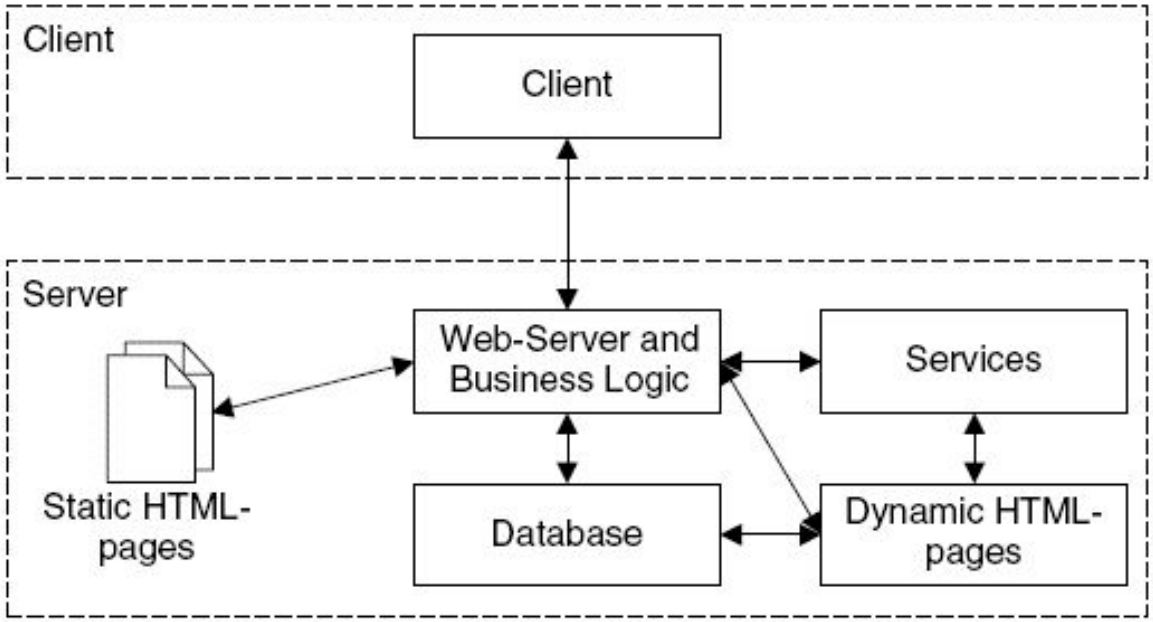
- HTML elements as **objects**
- **Properties** for all HTML elements
- **Methods** for all HTML elements
- **Events** for all HTML elements

The **HTML DOM** is an API for JavaScript:

- JavaScript can add/change/remove HTML elements
- JavaScript can add/change/remove HTML attributes
- JavaScript can add/change/remove CSS styles
- JavaScript can react to HTML events
- JavaScript can add/change/remove HTML events



Web Applications Architecture



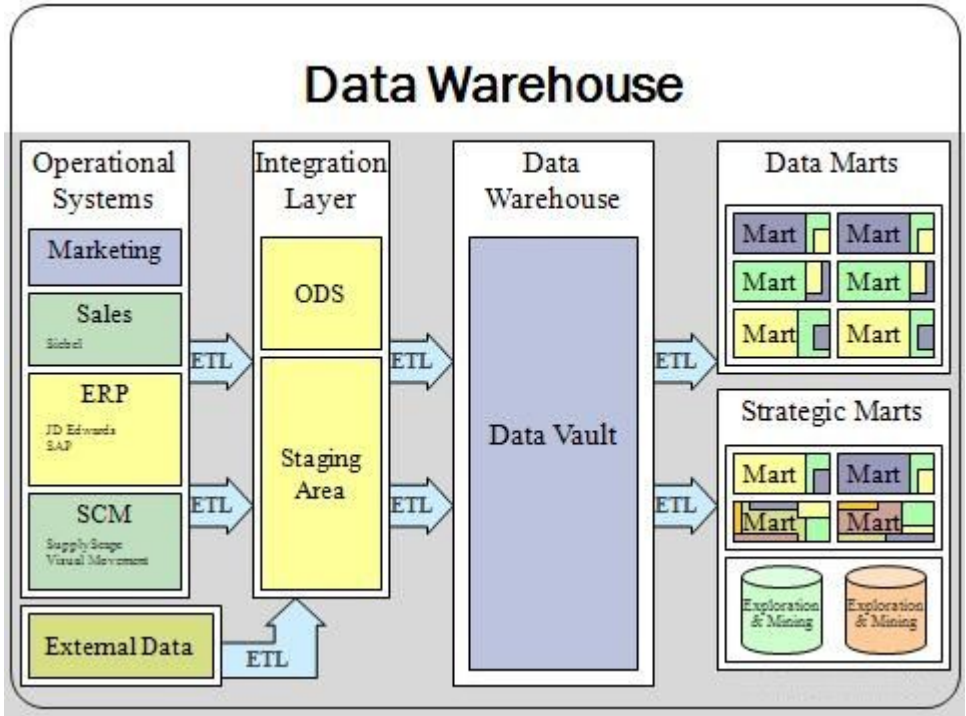
Data Persistence for Analytics

upf. Universitat Pompeu Fabra Barcelona

Data Persistence for Analytics

Data Warehouse (DW)

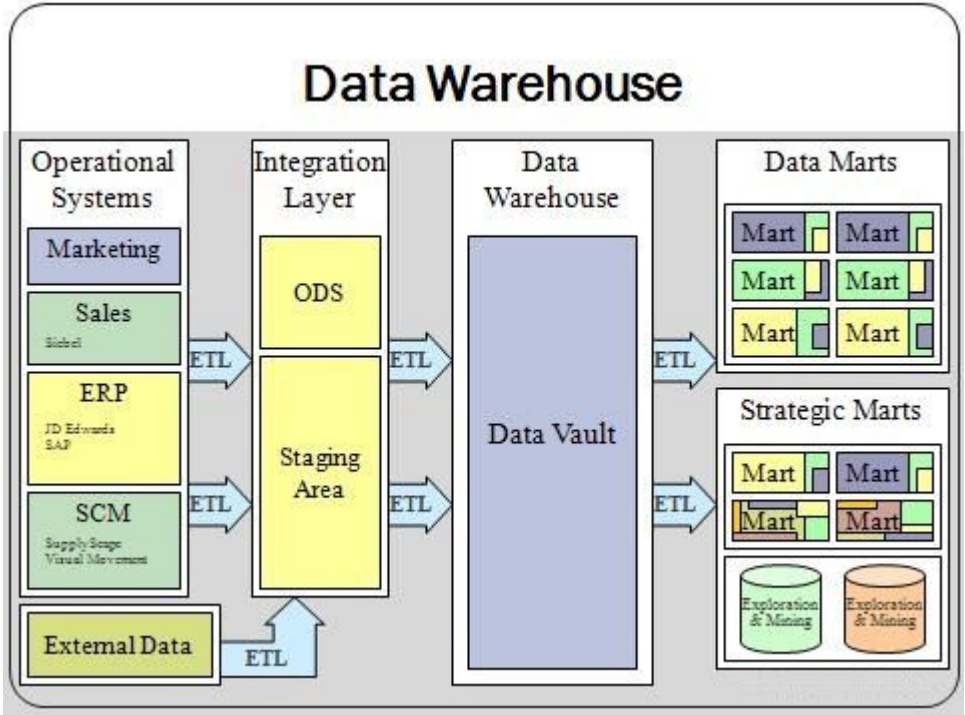
- Big Databases
- Holds multiple subject areas
- Holds very detailed information
- Works to integrate all data sources
- Does not necessarily use a dimensional model but feeds dimensional models.



Data Persistence for Analytics

Data Marts

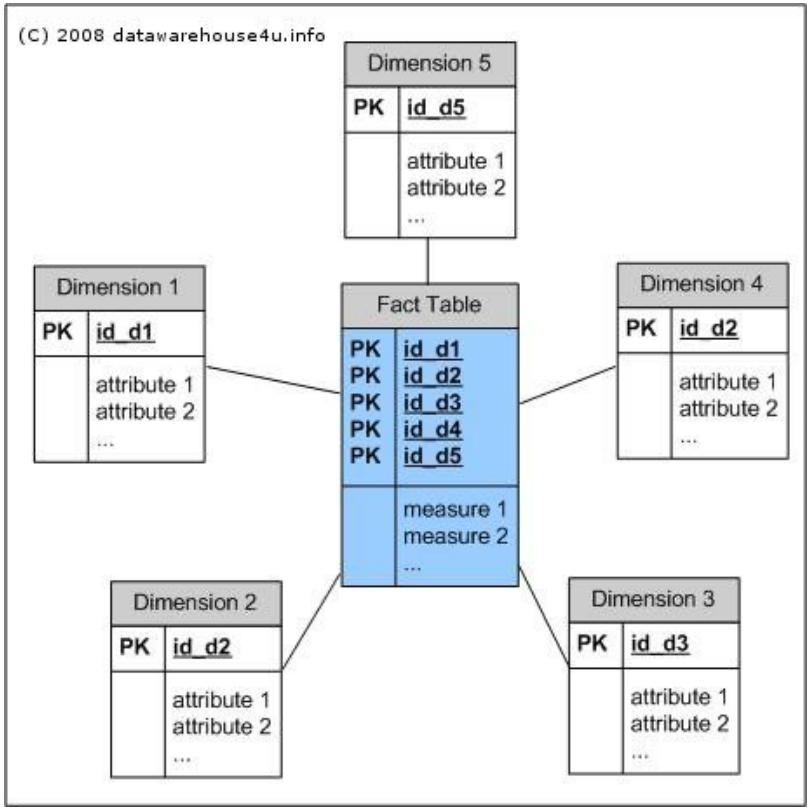
- Holds only **one subject area,** for example, Finance, or Sales, **or Web Analytics!**
- Holds more summarized data (although may hold full detail)
- Concentrates on integrating information from a given subject area or set of source systems
- Is built focused on a dimensional model using a star schema.



Data Persistence for Analytics

Star Schema

- Structure of database tables.
- Used approach to develop data warehouses and dimensional data marts.
- Usually have a large fact table surrounded by smaller dimension tables.
- Dimension tables do not change very much.
- Most of the information that the users need are in the fact tables.



Data Persistence for Analytics

Fact Tables:

- They record huge amounts of quantifiable business metrics for a specific event.
- **Transactions:** facts about a specific event (e.g., sales events)
- **Snapshots:** facts at a given point in time (e.g., account details at month end)
- **Accumulating snapshot tables:** aggregate facts at a given point in time (e.g., total **month-to-date** sales for a product)
- **Examples:**
  - Sales, expenses, inventory.
  - Sales price, sale quantity, and time.
  - Distance, speed and weight.
  - **Web usage statistics**.

Data Persistence for Analytics

Dimension Tables (DT):

- Contain static or descriptive data.
- Have small number of records, but may have a very large number of attributes to describe the fact.
- Examples:
  - **Time DT:** describe time at the lowest level of time granularity for which events are recorded in the star schema
  - **Geography DT:** describe location data, such as country, state, or city
  - **Product DT:** describe product models, colors, sizes.
  - **Employee DT:** describe employees, such as sales people names, contact information, etc.
  - **Range DT:** describe ranges of time, dollar/euros values or other measurable quantities to simplify reporting

Summary

Web Technology:

- HTTP protocol and client/server model.
- Browsers allow tracking web events.
- DOM HTML model and JS API.
- Web Server logic allows to persist statistical data

Summary

Data Persistence:

- Web usage data is huge.
- Use a star schema similar to Data Marts
- Facts for our web analytics topics.
- Dimensions for the facts

Web Analytics Features

Typical Features

- Monitor analytics via a dashboard.
- See heat maps of where users click.
- View detail usage of any web page including prior and next pages viewed.
- View the clickstream of user sessions.
- Define and track an unlimited number of website actions.
- Track which search terms are referring visitors to a given page.
- Analyzing how often users visit the website.
- Track clicks on all DOM elements on web pages.

upf.

Typical Features

- Tracking a Page View using Javascript
- Tracking a E-commerce Transaction
- Tracking Conversion Funnels
- See the Click Heatmap for a page
- See the DOMstream Recordings for a page
- Action Tracking

Action Tracking

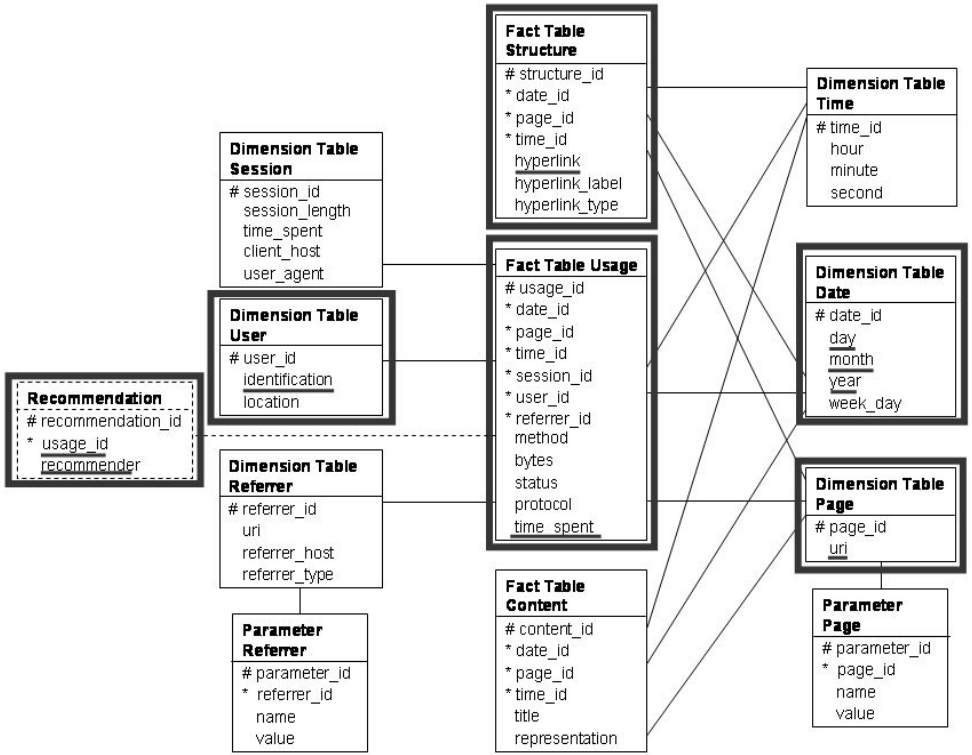
- Provides a way to track actions that users perform on your web site.
- An action can be anything:
  - a web form submit
  - a clicking on a particular link
  - A click on a particular UI control (button, dropdown, slider, scrollbar).
- Action tracking can be triggered from javascript or from your code.
- Actions are reported separately from pages views.
- Actions have their own metrics & dimensions.

Tracking Website Usage

Tracking Website Usage

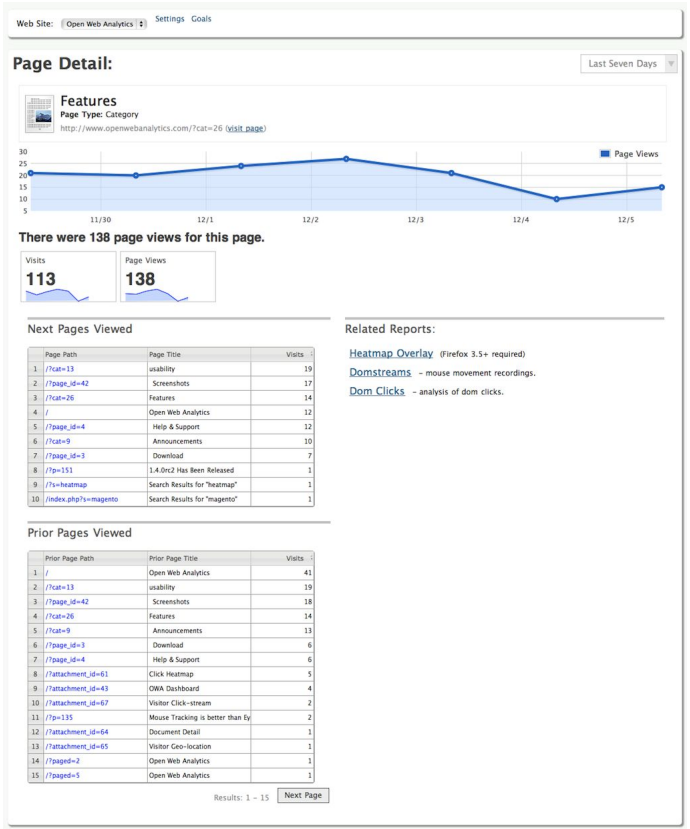
Data Collection:

- Define facts
- Define dimensions

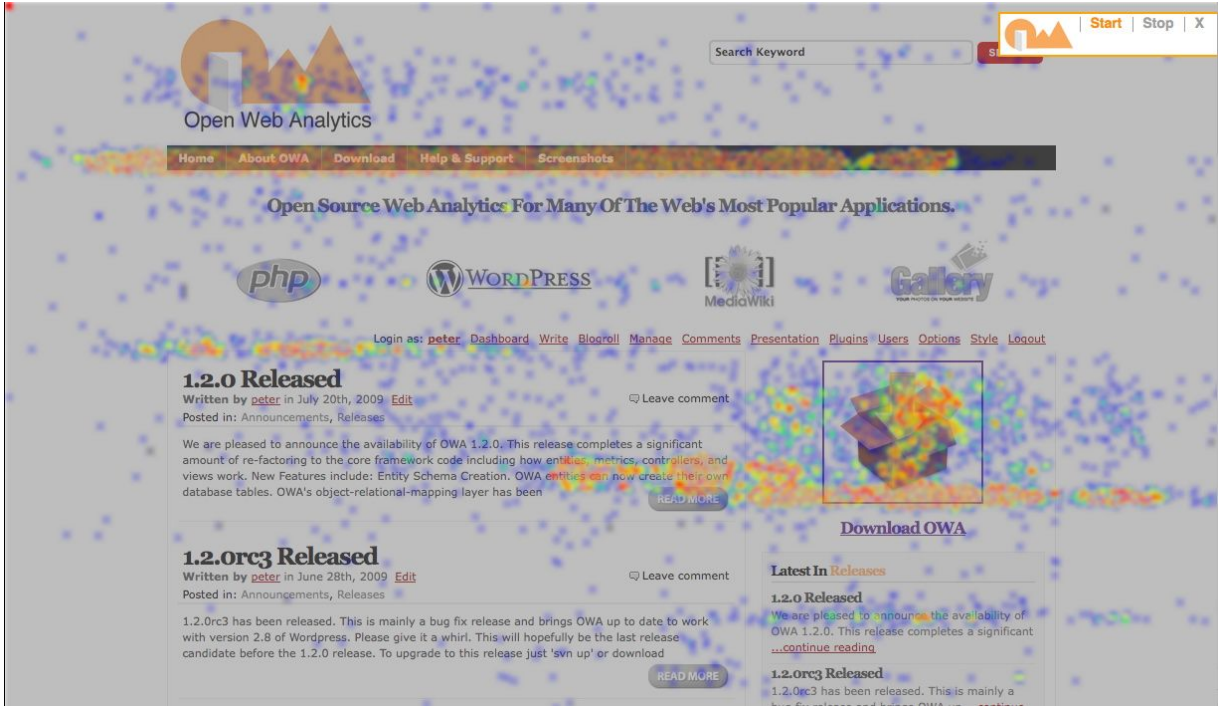


Displaying Usage Statistics

Displaying Usage Statistics

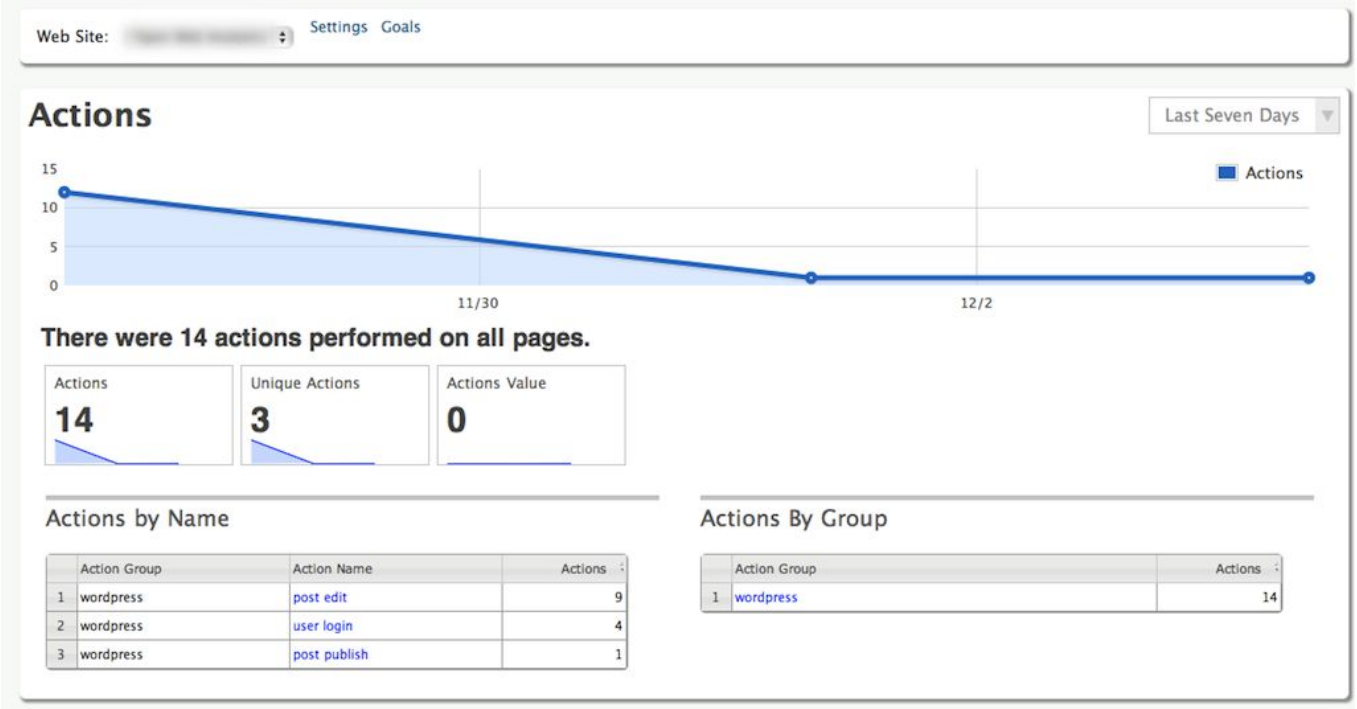


Displaying Usage Statistics



Mouse Clicks

Displaying Usage Statistics

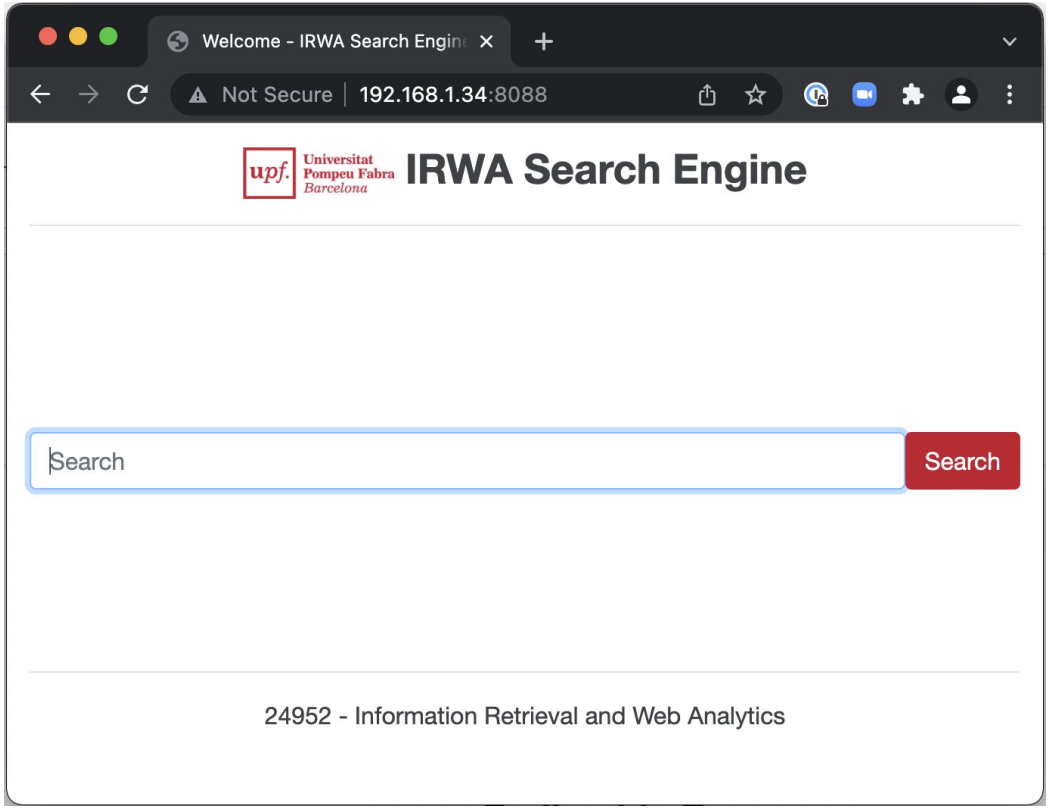


Actions

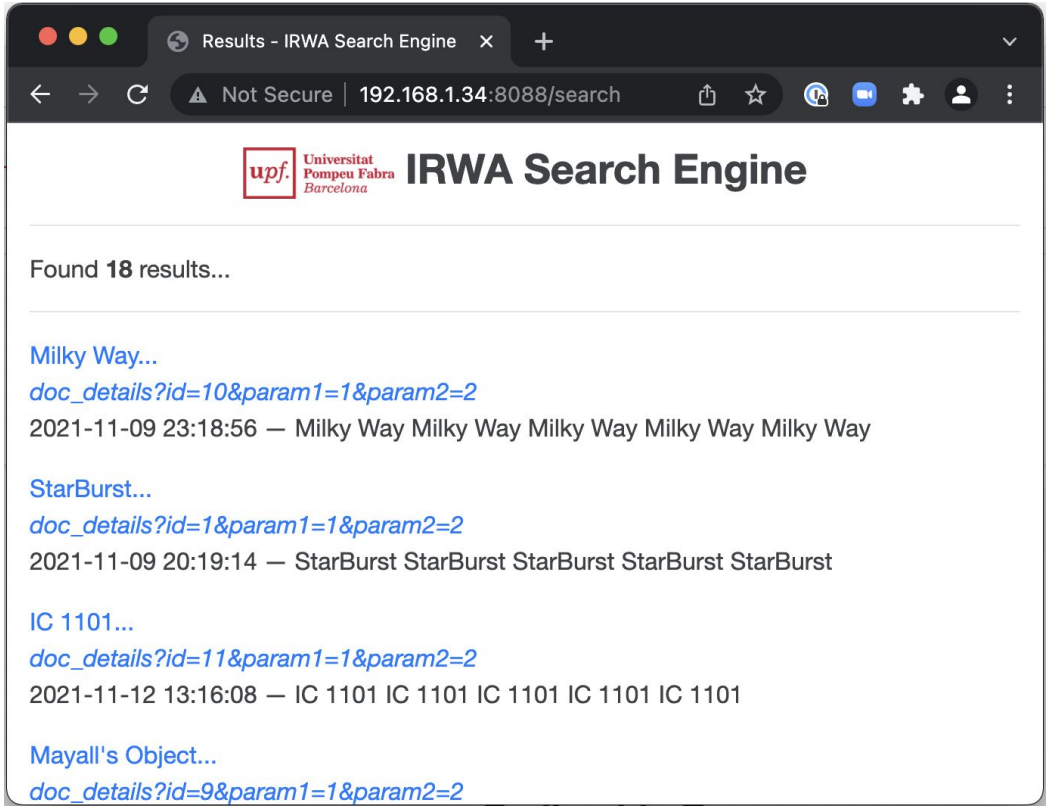
Implementing a Web Search Engine

upf. Universitat Pompeu Fabra Barcelona

Search Engine Web Application



Search Engine Web Application



upf.

Search Engine Web Application

- Flask framework (in Python).
- Simple web app structure.
- Jinja templates (HTML + logic code).
- HTTP request parameters.
- Using in memory data.
- Registering usage actions.
- Integrating search engine into the web app.
- Use objects!

Thank you!