John Pratt
CS 2340-A
Team 58
903046512

MediaWiki Individual Case Study

Introduction

The earliest iteration of MediaWiki was written by Magnus Manske (a university student at the time) in PHP to serve as a dedicated wiki engine for Wikipedia, starting in 2001. Manske intended to improve the performance of Wikipedia by converting it into a database-driven app, using an MySQL. The reason for this transition was that at the time Wikipedia was very simplistically implemented - it used a pre-existing GPL wiki engine called UseModWiki, written in Perl and backed by plaintext files which didn't even track changes with articles titled in CamelCase. This may seem overly simplistic, but at the time such implementations were extremely common, especially when looking back to the early stages of web development schemas in the late 90s and early 2000s. Because of the simplistic nature coupled with the growing popularity of Wikipedia at the time, Manske's solution aimed to provide enhanced features unique to Wikipedia while simultaneously improving performance by creating a separation between the site's databases and web servers. "The PHP Script," as it was referred to, was the first iteration of what would later become MediaWiki.

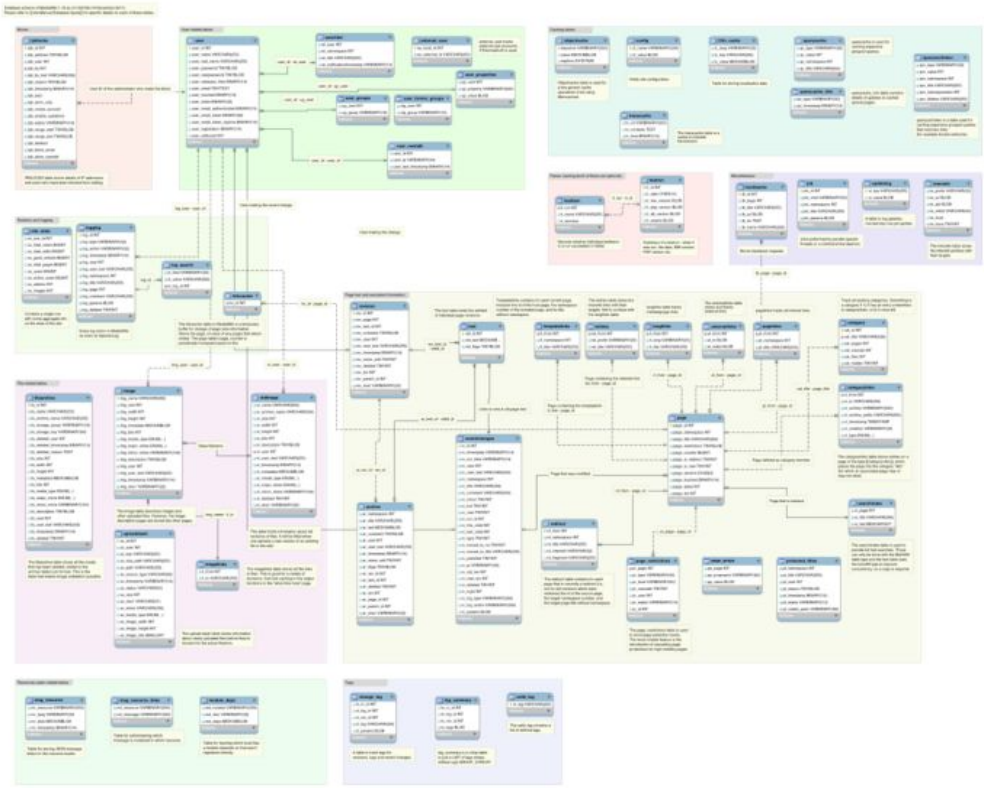
However, despite the improvements from the PHP script and having a database backend for the site, a culmination of limited funding for expensive features, limited hardware availability, and increased popularity continued to prove problematic for the site. In 2002, Lee Daniel Crocker rewrote the script, titling it "Phase III." It wasn't a complete refactoring of the PHP script, but rather meant to be a temporary hack to restructure existing code in order to streamline the site and increase efficiency. A larger problem began to form though as the site continued to increase in popularity - it needed to be able to balance its limited budget with enhanced security and increased customizability of the site. By 2003 this became a huge problem, as the developers realized that stomping out bugs as they came up was beginning to impact productivity - and possibly the future of the site as a whole. They came to the conclusion that they would rework and tweak the existing codebase moving forward instead of completely re-engineering a new architecture for the site, simply out of interest for time and the limited funds available for the project (that is, the entire project at this point had no full-time committed developers). By mid-2003 two major changes in the project occurred: Jimmy Wales (founder of Wikipedia) created the Wikimedia foundation (which became the basis for the naming of MediaWiki), and the first database server was added - thus beginning the project's large-scale application of single responsibility principle. By August 2003, MediaWiki had transformed into the earliest recognizable form of what we know today as Wikipedia - full scale editing; enhanced modern web customizability provided by javascript, HTML, and CSS; interwiki linking; maintenance and meta-discussion pages regarding the state of the site.

These factors combined are what made possible Wikipedia today - a free encyclopedia accessed and updated millions of times daily.

Description

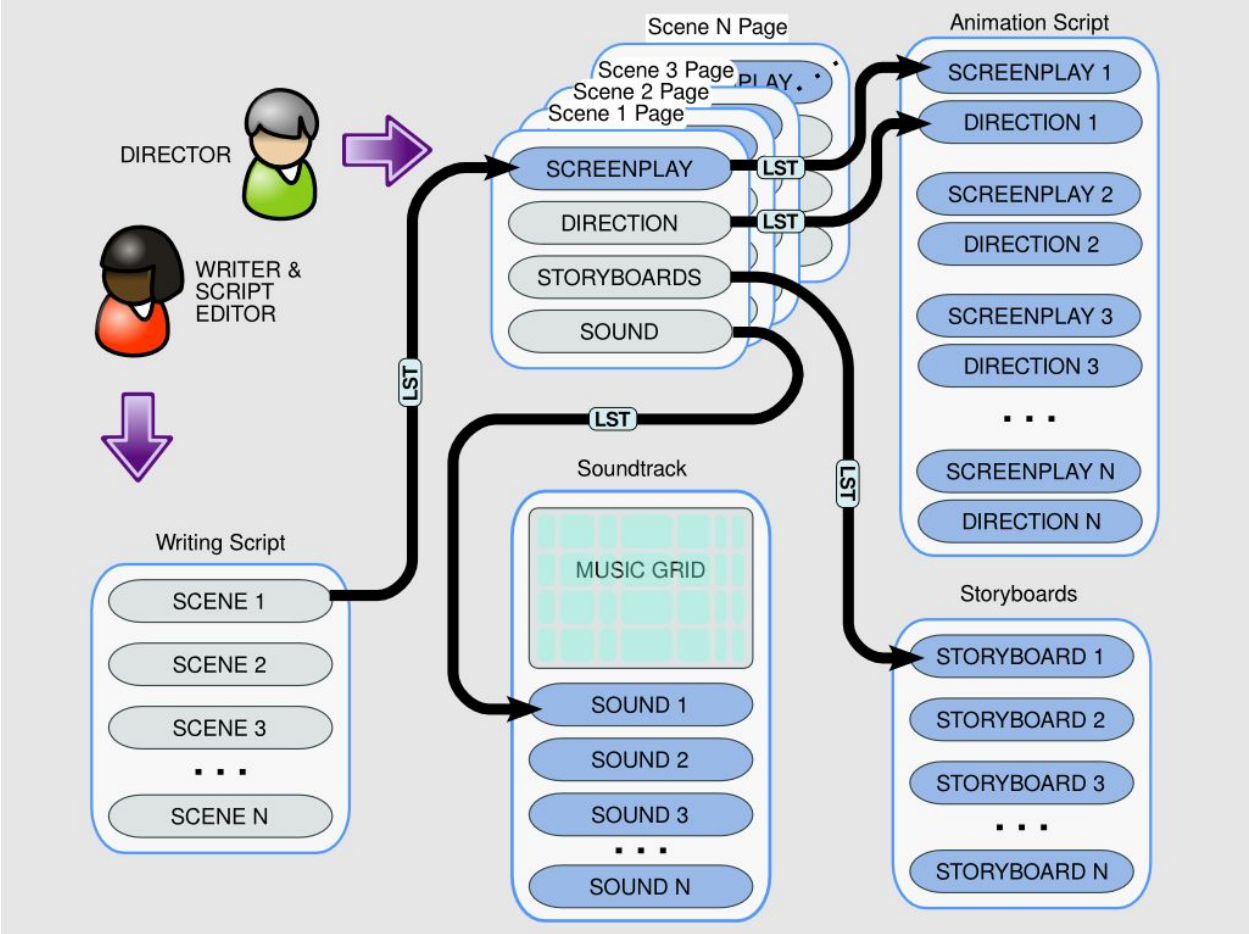
John Pratt
CS 2340-A
Team 58
903046512

For Wikipedia's Phase II software architecture, PHP was chosen. During the early years of development many developers didn't contributed well enough; However, in today's era there are many developers who are contributing their time and energy to make MediaWiki great. The originators of the MediaWiki didn't implement few abstracts, which is pretty perilous. Although PHP is more popular than Java, developers realize that if the MediaWiki was originally developed using Java then it would have been much better to code, more efficient, and easy to maintain the backend. Moreover, MediaWiki has been utilizing a relational database backend since the Phase II programming. The default database administration framework is MySQL, which is the one that all Wikimedia destinations use, however different DBMSes have group upheld usage. A sysadmin can pick a DBMS while introducing MediaWiki, and MediaWiki gives both a database reflection and a question deliberation layer that improve database access for engineers.



Even though MediaWiki contains inefficient code, since last past decades the major changes into code has been made and introduced new architectural elements. Later on, Parser, SpecialPage, Database, Image and FileRepo class hierarchy ResourceLoader, and the Action hierarchy was implemented, which supports the features which were designed originally. For easy of development, founder developers and reviews have enforced strict security parameters. To make it easier to write secure code, MediaWiki gives developers covers around HTML output and database queries to handle escaping. Then it scans the user inputs, goes thru the WebRequests class. WebRequest class evaluates the scanned data in URL or any other form. It does eliminates quotes and slashes, strips illegal input characters and regulates the Unicode

sequence. MediaWiki provides an XTML purifies with Sanitizer class and database functions that prevent SQL booster.



The current layout contains dozens of tables. Many are about the wiki's content other tables include data about users, media files, caching and internal tools. Indices and summary tables are used extensively in MediaWiki, since SQL queries that scan huge numbers of rows can be very expensive, particularly on Wikimedia sites.