



Difflet - Difference Between Entities

Monsoon 2016 Semester Project

05 Aug 2016

Project Guide: Prof. Radhika Mamidi

Project Members

Harshendra Avabratha [201505520]

Utsav Chokshi [201505581]

Md Tareque Khan [201505521]



Table Of Contents

[Overview](#)

[Present Solution](#)

[Project Description](#)

[Scope](#)

[Tools/Language](#)

[Architecture Diagram](#)

[Back End Components](#)

[FrontEnd Components](#)

Overview

We often hit searches for “difference between” two entities. When we want to know the differences in various features or properties in terms of numbers or definitions or other forms.



Example: C vs C++, India vs Australia, Panther vs Leopard

Present Solution

As of now there is no solid web application exclusively, which gives differences between two entities in a proper point-by-point format. Usually we have to look in to Q&A sites like Stackoverflow, Yahoo Answers and Quora.

Project Description

So we present an idea of a webapp which is similar to Q&A site where a user can view differences between two entities. The differences are extracted using Information Retrieval techniques. The similarity between the entities is automatically detected and reported if they are dissimilar. If differences are not present in the database, it will be created on fly and will be updated later.

Scope

1. Given two entities under a category display point by point difference.
2. Disambiguation of user input and also of the crawled data.
3. Fast retrieval of differences on the fly.
4. Create a web application for the user.

Extended Scope: Difference summary on same sections from wikipedia

Tools/Language

Python and Java: For creating the wikipedia indexer and Data Extractor Tool

Web2py Framework: for the front end web app

bash scripting: to help with different functions like integration and process start/stop

Architecture Diagram

Back End Components

1. Data extractor from one page
2. Wikipedia Indexer
3. Difflet generator

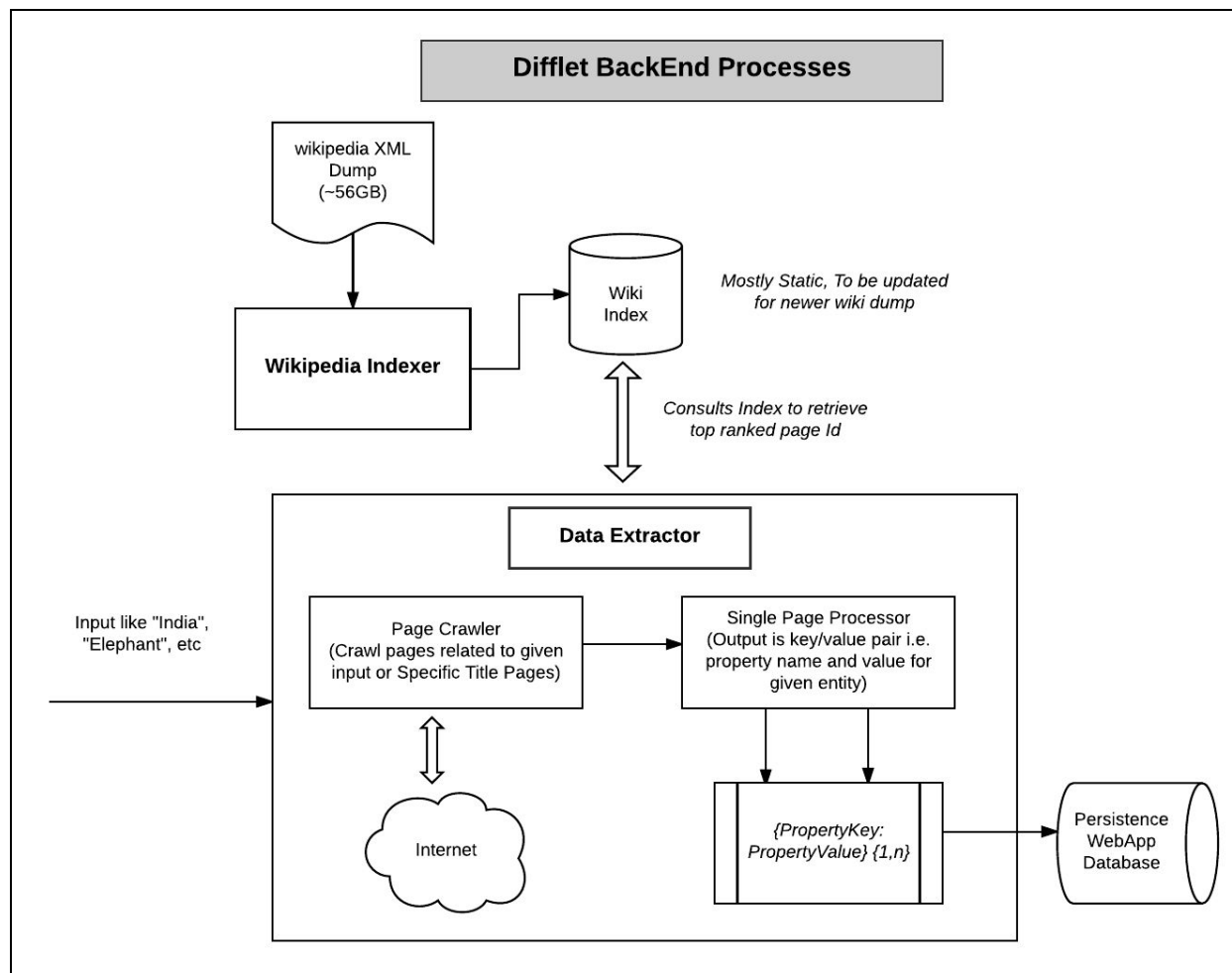


Fig 1. Difflet Backend Architecture

FrontEnd Components

1. Webpy webapp
2. Persistence Database

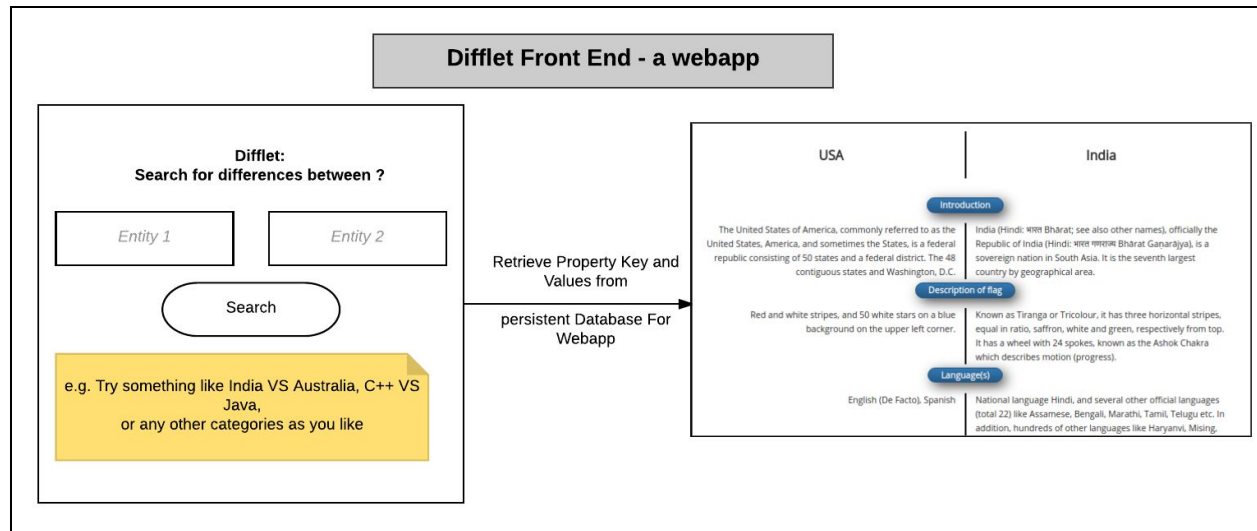


Fig 2. Difflet FrontEnd Architecture

Milestones

1. Working end to end model
2. Extraction of points on the fly for any given entity. [category to be extracted automatically]
3. Final Output with working website