

**Project Status Report June '16**  
**'Difference Between' Point by Point difference generator**

Team Members:  
Harshendra Avabratha  
Utsav Choksi  
Tareque Md Khan

**Work Done in two aspects:**

1. We have dry run the Wikipedia data indexer to get index for mini dataset dump of wikipedia. Querying for any entity e.g. India, Java gives the page-id of top related pages.

We tried to build a set of properties to extract from wikipedia pages (currently pure text saved into files), using text processing, but not have any better results.

E.g. Property set for countries comprises of (population, area, national anthem, languages spoken, capital, location, currency, etc)

For Programming languages, the property set would include ( creator, typing discipline (static/dynamic), license, Object Oriented/Functional, regex support, etc)

2. User Interface: We thought of creating a website alongside. We are using web2py, html5 and css to build a simple interface where user can choose a category and enter two names of entities in that category.  
We are thinking of adding user interaction to edit/update/add data for entities.

Link to UI: <https://mdtareque.pythonanywhere.com/difflet/default/search>

Example: C vs C++

<https://mdtareque.pythonanywhere.com/difflet/default/difflet?e1=c&e2=c%2B%2B>