

Problem outline and defining expected outcome

Study the role of Wikidata in enhancing the productivity of the Wikipedia writers and enhance Wikidata for Hindi and Telugu using gamification.

Name - Himanshu Maheshwari
Roll No - 20171033

Mentors - Tushar Abhishek and Nikhil Pattiisapu

The problem statement seeks to enhance Hindi or Telugu wikidata. Each wikidata entry requires a reference. any webpage on the internet could be used as a potential reference for a wikidata entry. Now about 56% of the webpages are written in English while only 0.1% of the webpages are written in Hindi. This stats raises a serious concern which is lack of credible Hindi or Telugu sources which could provide data that could be used to enhance Hindi and Telugu Wikidata, which could be used to increase productivity of Wikipedia writers.

Now to tackle the problem of lack of credible Hindi or Telugu sources we decided to use news articles from various renowned news outlets. The advantages of using news articles is that since they are written by renowned news outlets they carry some accountability and factual correctness and thus could be used as references. For the scope of this project we only focus on news relating to cricket. Thus the overall aim of the project is to develop a system that would enrich Hindi or Telugu cricket related entries using gamification.

Problem Outline

1. What is the problem?

I would define the problem as enhancing Hindi and Telugu Wikidata using gamification and then using that wikidata to enhance productivity of Wikipedia writers. Thus the first part of the solution would aim at enhancing Hindi and Telugu Wikidata and the second part will aim at providing users with this data so that they could use it to enhance wikidata. A common solution will achieve both the aims.

2. Difficulties and its solution

As already mentioned in the starting that the amount of web pages in Hindi or Telugu are very less and thus accountable and factually correct source of information in these languages are very less. Thus we aim to tackle this problem by using news articles from renowned news outlets as they bring in some accountability and correctness. For the scope of this project we only focus on news relating to cricket.

3. Solution (Defined in great detail later)

The solution/game could be divided into two subsystems:

- a. The first subsystem scrape news articles from various credible news outlets and does appropriate cleaning. After that it classifies the news article into one of many categories. The user is then asked to play for any one of these categories.
- b. The actual game
There are two types of players. First kind of players, called Beta players, check if the articles have been correctly classified or not. This helps in improving the classifier. The second kind of players, called alpha players, helps in extracting relevant information from the articles in a multiplayer scenario. There would be incentives urging them to play fairly. Finally when we get facts from Alpha players, we make Wikidata edits.

Expected Outcome

The outcome of the project would be a two subsystem game which would be able to:

1. Scrap new articles from various credible news outlets.
2. Classify news into one of many categories.
3. Check and improve this classification.
4. Extract relevant information from the articles in a multiplayer scenario.
5. Make Wikidata edits.

Solution Outline - First Draft

As already discussed the game/solution is a two subsystem game defined below:

Environment Creator

As the name suggests, the aim of this subsystem is to create the environment for the game. By environment I mean new articles. Now firstly there is a scrapper which scraps news articles from credible news outlets. Basic tools like beautiful soup etc. should be sufficient for this. As of now I am considering following news websites:

- a. Rajasthan Patrika
- b. Dainik Bhaskar
- c. NDTV news
- d. Hindustan Live
- e. Aajtak
- f. Zee News

Reason for choosing them is at least for cricket news these outlets could be trusted and they have a special page dedicated just for cricket.

After getting the news articles they would be categorized into following categories based on what they talk about:

- a. Players

- b. Stadium
- c. Records/Stats
- d. Country
- e. IPL
- f. Leagues other than IPL
- g. Domestic Cricket
- h. Umpires
- i. Cricketing Accessories
- j. Cricket Boards
- k. Miscellaneous

The Game

There are two categories of players:

1. **Beta Players** - Initially all the players are beta players. The aim of the beta players is to verify whether the articles have been correctly classified. They are upgraded if they have verified say 50 articles.
2. **Alpha Players** - These players compete in a multiplayer environment to identify lines containing facts that could be used to update wikidata.

If an article gets more than the threshold number of negative reviews from Beta players, its category is updated accordingly.

Alpha players could decide to play in any category they want to. A certain number of alpha players, say 10 will play the game together in an online multiplayer scenario. These players will initially be given a guide/tour/recommendation on how to find facts from articles. These players do not know each other. They all will be given articles from the category which they choose and they have to find facts from the article which could be used to update Wikidata. If a certain fact gets flagged by more than the threshold number of players (say 5) then it is used to update Wikidata and the link to the article is used as a reference. The player that extracts the most number of flagged facts wins. Players finally get points based on their performance. Since the players do not know each other and everyone wants to get maximum points, by game theory, they are only left with only one option i.e. play honestly. We might penalize someone for wrongly marking a fact. Also there will be rewards for players that cross certain points like in mobile app called Feetapart. In Feetapart the users are incentivised to walk by giving them coins if they have walked certain steps. These coins could be used later to claim rewards like Cult fit membership, Oats etc. Similar rewarding system could be used over here also. After getting the facts we can make Wikidata edits (right now I am thinking of manually doing so but later will automate that).

Use Case Diagram

