Using Machine Learning to Automatically Score the Authoritativeness of Wikipedia Articles

Chris Salij
Online Dublin Computing Science Summer School (ODCSSS),
Clique Research Cluster, UCD

***Abstract*—Creating and maintaining encyclopedic quality content is the central goal of Wikipedia. However at the time of writing, there are over 3.3 million articles on Wikipedia, many of which are not at the standard that Wikipedia aspires to. Considering the sheer volume, even identifying the pages which need work, from the ones that do not is a monumental task, and would require a significant amount of effort for even this preliminary step. In this paper we present my research into creating an algorithm for automated page classification which would help in the aforementioned situation. We take a user centric approach in classifying pages, where the quality of a page is judged based on the users that have contributed to it, and how much they have contributed. We use supervised machine learning to classify users based on how reliable they are to produce good quality content in a given domain. We then added that classification to a labelled data set with key attribute about a page, on which we again used a machine learning algorithm to identify good and bad quality pages.**

***Index Terms*—Web 2.0, Wikipedia, Machine Learning, Classification…**

I. Introduction

The goal of this project was to create an algorithm which can automatically determine the quality of a page, and assign it a rank. Wikipedia already has a number systems in place to rank the quality of articles from high quality (Featured Articles) to low quality (Stub/Start). However the existing methods are all highly manual processes which require a significant amount of volunteer time and effort (see Section 2.1). If the classification of pages could be somewhat automated, then the time and effort that is spent classifying pages, could instead be put to work improving the pages themselves. In other words, reducing the amount of administration that users have to do in order to help improve Wikipedia.

Our approach to solving this problem is to take a user centric analysis. There are several reasons for this, the main one being that Wikipedia is a user centric organisation itself. All the work on Wikipedia comes from users who volunteer their time and effort to make Wikipedia better. From the content itself, to managing the community, it is all done by volunteer users. There are a handful of people who are paid by Wikimedia Foundation, but they are mainly engineers who look after the servers and technical aspects of the site. Given Wikipedia's user focus, then it seems that one of the best ways to judge the quality of content is based on the users themselves. Their actions on the site, as well as their interactions with other users make the site what it is. We believe that it is through this scope that the best judgement about the content on Wikipedia can be made.

II. Background

Wikipedia was launched in 2001 and is a user created and maintained Web-based encyclopedia. It differs from traditional encyclopedias in that its content does not necessarily come from experts in a given field, but from anyone who wishes to contribute to it. The idea behind letting anyone edit an article on Wikipedia, is that the more people that view an article, the more likely they are to see a mistake. Thus they can correct that mistake and the encyclopedia becomes better. This is the "Many Eyes" idea, kissing-cousin to the "Many Hands" (1) idea.

Wikipedia's approach of using the "Wisdom of the Crowds" to produce its content is controversial, with many advocates for and against the practice (see Korfiatis et al)[2]. Some claim that the content it produces is equal to that of traditional Encyclopedias while others strongly disagree. Regardless of the controversy, Wikipedia is currently the largest and most popular general reference work on the Internet.

While Wikipedia contains many high quality articles which have been heavily copy-edited and well sourced, and are on-par with, or perhaps better than traditional Encyclopedic equivalents. There are also many articles which have not received as much attention, they languish with sub par content, opinionated viewpoints and few-to-no sources for their claims. In articles such as these vandalism and downright false information is all too common. In essence, they fall short of the minimum requirements of being an encyclopedia article and cannot be trusted. This is where there are a number of efforts have been started to improve the content on Wikipedia.

1. http://dictionary.reference.com/browse/Many+hands+make+light+wor

A. Existing Methods

The main method for ranking and improving content on Wikipedia are WikiProjects(1), in which a group of users get together and pick a Category (almost all pages on Wikipedia are organised into one or more categories)(2). This group of users then proceeds to analyse and rank all the articles in that category according a scale (fig 1) from FA (Featured Article) to Stub. (There are other assignments available, but they don't pertain to quality). This information is then used as something similar to a to-do list, where people who wish to help improve Wikipedia are directed toward pages that are identified as poor in order to improve them.

However, as described in the introduction, this method is extremely laborious. It requires that multiple users read over each article in the category and agree on a ranking to assign to it. In the case of the History category (see Section 3.1) there are over 15,000 articles. As we can see in fig 1, just over half of those articles have been rated according to the scale. That leaves 6,500 articles still to be assessed, a monumental challenge in itself. That is just one sample WikiProject, and there are dozens of Wikiprojects, many with overlapping pages, and yet the case is similar for almost all of them. A large proportion of pages are unranked, and attention is not brought to them until someone has the time to manually read through the page and rank it.

While this method is in keeping with the spirit and nature of Wikipedia it is obviously a very inefficient system.

History articles by quality and importance								
	Importance							
Quality	Top	High	Mid	Low	NA	Other	???	Total
FA		2	11	8		106		127
FL				1		2		3
A			1			1		2
GA	2	2	11	15		176	16	222
B	10	23	41	26		564	92	756
C	19	40	39	31		406	36	571
Start	14	33	72	146	2	2,598	330	3,195
Stub		4	24	256		2,583	234	3,101
List		13	64	6	6	39	10	138
Category				1	205	57		263
Disambig				1	7	4		12
File					3	5		8
Portal					3	11		14
Project					2	9		11
Template	1	1	1	1	38	35		77
NA					9	99		108
Assessed	46	118	264	492	275	6,695	718	8,608
Unassessed					3	5,509	992	6,504
Total	46	118	264	492	278	12,204	1,710	15,112

(Fig 1)[4]

1. http://en.wikipedia.org/wiki/Wikipedia:WikiProject
2. http://en.wikipedia.org/wiki/Wikipedia:Categorization

B. Similar Research

B.1 A Content-Driven Reputation System for the Wikipedia - Adler & Alfaro

Adler & Alfaro[1] conducted similarly focused research into assigning a user reputation based on the content of edits which that user had contributed and how those edits had survived over time. They looked at two main metrics:

- Text life - How much of a the text inserted by one user is still present after a subsequent edit.
- Edit life - How much of the reorganisation (text ordering and deletion) performed in one edit is still present after a subsequent edit.

From these two metrics, as well as edit distance[5] they assign each user a reputation on a per contribution basis. Thus they build a picture over time of a users real contributions and how those contributions have been recieved by the Wikipedia community. In their research they analysed the entire Italian Wikipedia site with 154,621 articles holding 714,280 filtered* revisions, as well as the entire French Wikipedia site with 536,930 articles and holding 4,837,243 filtered* revisions. In their analysis they found that their approach had a reasonable correlation between what it predicted that the user would produce and how that users contributions actually fared over subsequent edits.

Their research differs from the research layed out in this paper in that Adler & Alfaro analysed, in-depth, each of a users contributions and how it survived over time. They then assigned a rank based on that analysis. However in this paper We have looked at user data in aggregate, without analysing each edit individually. For how Adler & Alfaro's work may be directly incorporated into the work layed out here see Section 5.2.

* Adler & Alfaro filtered edits where the same user had multiple sequential edits. In their analysis they only considered the final edit in a string of edits by the same user, and discarded the intermediate edits.

B.2 Evaluating Authoritative Sources using Social Networks: an insight from Wikipedia - Korfiatis, Poulos & Bokos

Korfiatis, Poulos & Bokos[2] analysed user interactions by computing the contributor degree centrality and article degree centrality of each user and article respectively. Their research attempted to uncover how author interaction and distribution may be a key factor in producing high quality Wikipedia articles. Their analysis did not look at the individual edits of users and how they interacted on a small scale, but it analysed the cumulative figures and modelled their interactions.

One potential downfall with their research, which they address in their paper, is that their methods would only ever produce figures for community consensus and not universal truth. However this will always be true when you are analysing users themselves rather than the content directly. But it is still a salient point to make, that no matter what the user analysis may be, if the community or a significant part of the community is biased in one direction and the content is slanted, then measuring community consenus will never be a measure of truth. For further analysis see Section 6.

B.3 Us vs. Them: Understanding Social Dynamics in Wikipedia with Revert Graph Visualizations - Suh, Chi, Pendleton & Kittur

Suh, Chi, Pendleton & Kittur's[3] research was focused on visualising reverts, and using reverts as a method for modelling user conflict in articles on Wikipedia. In their work Suh et al assumed that reverts are proxies for disputes and disagreements, and in analysing a number of controvertial articles with high numbers of reverts, they classified users visually based on their reverts. They grouped users together using a force directed layout graph in order to simulate social structures. The resulting groups of users then tended to share common viewpoints on a given controvertial article or topic. Thus through their analysis they managed to roughly determine the bias of users editing a given Wikipedia article, as well as users which did not exhibit any bias in their reverts.

Suh et al faced similar problems as we did (as described in Sectionn 5.1) in attributing reverts in articles where there we large numbers of reverts with many people editing over one another. However they simplified their problem by only counting the revert relationship between the reverter and the editor who made the immediate last edit regardless of any ambiguity. We decided that this particular approach was not accurate enough for our intentions and so we kept track of every user between the reverter and the edit to which it was reverting to. This made our solution slightly more complex but we felt the solution was worth it.

III. Project Work

In this section we outline our work in the project and how it was carried out.

First a note on the license the Wikipedia uses for its content. All content on Wikipedia is either Creative Commons Attribution-sharealike 3.0 Unported License (CC-BY-SA)(1) and the GNU Free Documentation License (GFDL), which expressly allows the work we have carried out in this paper. (2)

1. http://creativecommons.org/licenses/by-sa/3.0/
2. http://en.wikipedia.org/wiki/Wikipedia:Copyrights

A. Identifying a sample set

Initially we had planned on using two sets of sample pages, vandalised articles(1) and featured articles as our initial assumption was that Featured Articles would represent top quality pages and vandalised articles would inherintely represent low quality pages. However after some reserach into the matter it quickly became obvious that, while Featured Articles were always of a high standard, vandalised articles did not strictly mean low quality articles. In general, highly vandalism actually correlated with higher quality articles. We presume that this was because more people were involving in undoing the vandalism, there was therefore more attention being put on the article in general, and as people were undoing the vandalism they were also improving the content bit by bit. Another possibilty is that more famous or topical articles are more prone to vandalism, and those articles by their nature are under more focus due to their popularity and thus tend to be of better quality.

Whatever the reason, our finding made us rethink our initial assumption of Good articles vs Vandalised articles. We soon decided to use the rating system of Wiki Projects themselves (fig 1), as ranking articles according to this rating system is one of the final goals of this project.

Another aspect of this project, as outlined briefly in the description, is that we seperate user contributions in different domains. For example, a user may produce excellent and high quality content in one domain such as Medicine as they are an expert in that field. However they should not be given the same weight in a different field such as Engineering or History. Users should establish their compentency in each area before being assumed to produce reliable content. Thus for our sample set we decided to take a single category and work solely from that. We identified the History category as having a large number of rated articles to work off and thus we chose that as our sample set.

1. http://en.wikipedia.org/wiki/Wikipedia:Most_Vandalised_Pages

B. Gathering the Data

All the code for this project was done in Ruby. The fact that there was a strict 10 week time limit on the internship, combined with the ease of use and versatility of Ruby, made it the ideal choice. It facilites short, concise code that is simple to create and easy to understand. This allowed rapid prototyping and development of code along the way. One major downside of Ruby is the runtime speed due to the fact that Ruby is in interpreted language. In general interpreted languages run slower than compiled langauges. However the complexity of writing code which carried out the same function, in a faster language such as Java or C, totally outweighed any speed benefits that may have been garnered. We also had a server which we could leave running overnight, or over a weekend whenever a large batch processng job needed to be done, which mitigated the slower speed of Ruby.

Once we had identifed our sample set of pages (see Section 3.1) we needed a way to gather the data for those pages. We created a script which queries the WikiMedia API (1), as there was no existing one in Ruby which met our needs. This script takes in a list of pages and systematically fetches the last X number of revisions, and all the relevant data for those pages. The Wikimedia API provides a quick and simple, well-documented interface which was expressly created for uses such as this.

We also registered a Bot account(2) on Wikipedia which gave us the power to get more results in a single query. This was a major improvement to our gathering code, which minimised the number of queries that we had to make, and thus significantly decreased the run-time of the code. Registering a bot with Wikipedia provides a number of benefits, such as increased limits on a number of API calls. Wikipedia imposes lower limits on regular users so as to prevent automated spamming.

1. http://www.mediawiki.org/wiki/API
2. http://en.wikipedia.org/wiki/User:ChrisSalij_Bot

C. Storing the Data

The data that was downloaded using the aforementioned script was stored in XML files. There are a number of reasons for this, but the main one is that the data returned from quering the Wikimedia API comes in XML format. Thus the easiest way to handle it was to write it to a file directly and then deal with the contents at a later time. (The Wikimedia API allows a number of other flat-file formats)(1)

Once all the data had been gathered, then it was reorganised into an easier to analyse XML layout, where each user had their own file with all of that user's pertinent information contained in that one location. This meant that it was much easier to extract the data relevant to a user when it came time to analyse each user directly, rather than processing all the data we had gathered each time we need some specific information about a single user. We also had a similar layout for pages, where each page had its own small file containing all the pertinent information for that page.

We chose this method instead of setting up a SQL database or using sqlite, simply for the versatility that XML offers. It allowed rapid changing of the data stored in the files as our needs morphed and changed over the life of the project. On a number of occasions it was realised that we did not need to store one piece of data, or we needed to store another, and XML allowed us to change rapidly. Using SQL or sqlite would have added a layer of administration that was not needed.

1. http://www.mediawiki.org/wiki/API:Data_formats#Output

D. Feature Extraction

D.1 Reverts

For both users and pages, we needed to extract a number of attributes (see Section 4.1.1 & 4.2.1). Some of the key attributes we extracted were revert counts.

- A revert is when one user completely undoes a previous user's revision, and puts the article back to a previous state.(1)
- A user is reverted to if another user reverts an article to their edit (previous state)
- A user is reverted over if another user completely undoes that users revision.

Reverts are generally done in the case of vandalism, a negative edit, or an edit which adds nothing of value to an article. Thus revert counts were hoped to be a good indicator of a users reliability in the sense that if a user regularly had their own edits reverted over, then they were likely producing bad content. If a user has been reverted to, then it is essentially a vote for the quality of that edit. Similarly if a user had performed many reverts themself then it shows that they are getting involved reverting bad or vandalised content.

So determining if a revert has occured, who performed it, which user made the original edit, and which users had their edits removed in the process was a key goal. Since a revert is basically an edit which changes the content of an article back to a previous state, we simply hashed the text of each revision and checked to see if any previous revision matched that hash. The earliest occurance of a revision with that hash was the originating edit which was being reverted to. Each revision between the first occurance of the hash and the last were intermediate revisions (reverted over). Thus we were able to closely approximate figures for each user as to how many times they had reverted an article, been reverted over, or been reverted to.

1. http://en.wikipedia.org/wiki/Help:Reverting

D.2 Multiple sequential user edits

We discarded user edits where a user had multiple sequential edits in a row. If a user had a series of edits on the same article which were unbroken by other users editing in the meantime, then we only took the final edit. The main reason for this was to not inflate a users edit count, however it can be said that if a user has a string of sequential edits, then each edit is only an intermediate step to the final edit. This is the same action that Adler & Alfaro took in their Content Based Analysis of Wikipedia (see Section 2.2.1)

IV. Results

A. Classifying Users

We used the data mining software Weka(1) to help classify users. Weka has a large number of built-in machine learning algorithms. We trained it on a manually labelled dataset, in which we had manually classified 138 users into two groups, 'reliable' and 'other'.

- 'Reliable' users are defined as having a history of producing high quality content, and contributing positively to Wikipedia. These contributions may be reverting bad or vandalism edits etc.
- 'Other' users are defined as having a history of poor quality contributions or having not contributed enough to make a reasonable prediction as to the quality their future contributions.

In evaluating users manually we only considered actions within the group of articles we were classifying. This was to keep the data as precise and acccurate as possible and to prevent actions from outside our sample set of articles from contaminating our dataset.

We decided not to distingush between unreliable and new users due to the fact that, if new users are automatically given a higher ranking than known bad or vandal users, then there is nothing to stop an ill intentioned user from signing up for mulitiple 'puppet' accounts(2) and benefiting from the higher reliability level which a new user account brings. In this manner we hope to avoid gaming the system.

We also only considered registered Wikipedia users in our classification process. While unregistered users can freely edit most articles on Wikipedia, their edits are only attributed to the IP address which they used to make the edit. Since IP addresses of individual users tend to change over time, and they can also be changed easily by either rebooting a router (in the case of a Dynamic DNS), or simply moving to another internet connection. We decided that there was no point in evaluating anonymous users as their is no reliable way to attribute edits to a single user.

Initially we used Weka to see how internally consistant our manually labelled dataset was. Weka was over 94% correct in predicting 'reliable' vs 'other' users in a sample size of 138 using cross-validation and Logistic Regression. Given this high internal consistency in the training data, and the good predictive powers it had, we used this model to classify the other 13,000 registered users using the same model.

1. http://www.cs.waikato.ac.nz/~ml/weka/
2. http://en.wikipedia.org/wiki/Wikipedia:Sock_puppetry

A.1 Attributes

The following attributes were used as part of the users dataset

- **Bot:** Whether or not the user is a registered bot on Wikipedia
- **Edit Count:** The number of edits that a user has made on Wikipedia (Within our sample set of pages)
- **Page Count**: The number of pages that a user has edited on Wikipedia (Within our sample set of pages)
- **Reverted To:** The number of times a user has been reverted to, over another user.
- **Reverted Over:** The number of times a user has been reverted over, as in a later user has undone all of that user's edit

B. Classifying Pages

In classifying pages we used the data from the user classification that we performed above, as an aspect of a labelled dataset which we gave into Weka (see Section 4.2.1). We decided to perform a classification to see if the Weka could predict Good vs Poor pages based on the dataset we gave it (see Section 4.2.2) as this was in line with our goal of identifying poor pages which need work.

B.1 Attributes

These are the attributes we used in the labelled dataset which we gave to Weka.

Note: For the (type) columns below, anonymous users and users classified as ’other’ in the user classification (as described in Section 4.1) were added together to make a single figure which is represented as ’other’. This was due to the reasons outlined in Section 4.2 regarding keeping ’other’ users, unreliable users and anonymous users at the same level so as to prevent gaming the system etc.

- **Edit Count:** The number of edits (revisions) that page has had
- **Link Count:** The number of links (from pages within Wikipedia) pointing to that article
- **Revert Count:** The number of reverts which that page has had
- **User Count:** The total number of users that have contributed to that page
- **User (type) Count:** The number of users of each classification that have contributed to that page (reliable, other).
- **User (type) Percentage:** The percentage of total users of each classification that have contributed to that page (reliable, other).
- **User (type) Edit Percentage:** The percentage of the edits that each classification of user has made on that page (reliable, other).
- **Classification:** The classification that a Wikipedia users have given to the article according the the scale layed out in fig 1.

B.2 Dual Classification

In order to run this test we split up the ratings given by Wikipedia users into 3 groups using the ranks as seen in fig 1.

- **Good Pages:** FA (Featured Article), A & GA (Good Article) are the top three ratings and are all either encyclopedic quality content, or on the verge of it.
- **Mediocre Pages:** B & C are medium quality articles. Some basic work has been done on them, but they are not encyclopedic quality.
- **Poor Page:** Start & Stub. The miniumum amout of work has been done on these articles and they need to be greatly improved.

For our dual classification, we attempted to predict the quality of a page with only two classifications, Good vs Bad, using Weka. We decided to remove the middle group after some preliminary tests showed that 3 class classification when all three classses are on a continuum can lead to very muddled results. Given the fact that the classification of pages is a subjective task. Also the fact that the data we are analysing is not the critera for which a page is assigned a given ranking. Our task was to find a correlation between the data we have amassed and page quality. Both of these factors mean that when we introduce the middle class it leads to none of the 3 classes getting a firm definition for the quality of a page can be predicted against. Thus the predictions fail. See section 6 for more.

```
Correctly Classified Instances    212    70.1987 %
Incorrectly Classified Instances   90    29.8013 %

 a  b  c   <-- classified as
87 12  2 |  a = poor
33 41 27 |  b = medium
 2 14 84 |  c = good
```

(Fig 2)

Our dual class classification on the other hand, produced excellent results as can be seen in fig 3. We used the K* classification algorithm provided by Weka with a 10 fold cross validation on our labelled dataset. This resulted in a 97% correct prediction rate on the labelled dataset we supplied. It correctly classified 99 out of 101 pages as ’poor’ and 96% pages as ’good’. There was a 2 out of 101 false positive rate of ’poor’ pages classified as ’good’, and a 4% false positive rate for ’good’ pages classified as ’poor’.

```
Correctly Classified Instances      195      97.0149 %
Incorrectly Classified Instances      6       2.9851 %
Total Number of Instances           201

=== Confusion Matrix ===

  a  b   <-- classified as
 99  2 |  a = poor
  4 96 |  b = good
```

(Fig 3)

Unfortunately, further tests could not be run within the time constraints imposed by this internships.

V. Future Work

In this section we describe possible improvements to my work to date.

A. Revert attribution

One aspect of feature extraction which was very difficult to get working correctly was revert attribution. Detecting whether an edit is a revert, is a relatively simple thing to do, however when there are revert wars, with many users reverting over one another, it is difficult to attribute who is reverting over whom, and thus my method for calulating revert over produced figures that were slightly on the high side. Upon a manual inspection the figures from the algorithm tended to be about 10% too high for users with over 150 ’revert over’s. If these figures were to be made more accurate it could improve the predictive power of the approach used in this paper.

B. Individual edit analysis

We would have liked to incorporate an approach such as that taken by Adler & Alfaro[1]. In which they analysed the content of each edit and ascribed a reputation to the user based on how their edits survived over time. The scope of this project was looking at user actions in aggregate and classifying their likelyhood to produce good content based on that data. However we feel that if we had such granular content analysis we could greatly improve the predictive power of both the user classification as well as the page classification heuristic. In particular we feel that it would benefit the user classification since we are not directly analysing their edits in this paper.

C. User Interaction

One user feature which we did not model was user interaction. We had planned to model user interactions through things like collaboration through tandem editing, as well as disagreement and fighting through harsh edits and reverts. Unfortunately time did not allow for this to be done. It is my prediction that high quality articles would have high levels of user interactions, be they collaboration or fighting. Whereas low quality articles would have very low levels of user interaction and would have more fighting than collaboration.

Korfiatis, Poulos & Bokos[2] did a similar analysis in which they computed the contibutor degree centrality of a series of articles in an effort to identify the relationship between the collection of users who worked on an article. While this is not analysing how users interact directly, they did reach some interesting conclusions with analysing data in the aggregate.

D. Larger Dataset

Due to the time constraints imposed by the internship, our dataset was limited to what could be processed and analysed in a short amount of time. The next step would be to expand the dataset to all ~8,600 preclassified articles in the History category. If the prediction rate were to be close to that found in Section 4.2.2, then the following step would be to classify the remaining ~6,500 unclassified articles in the History category. Thus providing a real and tangable benefit of this research to Wikipedia.

VI. Summary

In this paper we have presented our method for scoring page authority using machine learning. Our analysis took a user centric focus, in which we classified a users likelihood to produce good and reliable content. We then took this analysis and used aggregated figures for user involvment in the editing of a page and added them to other key attributes of a page to form a labelled dataset. The labels came from the ranking that Wikipedia users had assigned to each page in WikiProjects.

We used this labelled dataset to successfully predict the quality of articles in a dual class classification senario in which the middle class had been removed in order to provide a clear distinction between the two outer classes. This test yielded a particularly good result rate of ~97% correct prediction rating, with very few false positive identifications on either side of the scale. One aspect of this result which should be payed attention to in particular, is that the false positive rate for 'good' pages, i.e. labelling a 'bad' page and as a 'good' page, was ~2%. This is a very encouraging result, as a tool like this would ideally be used on a large group of unranked pages in order to identify bad pages which need work. Producing a tool which can identify poor quality pages was one of the original goals. We feel that it is perhaps the most compelling outcome of this project and the one which could be most easily and readily utilised on Wikipedia the help alleviate the problems outlined at the beginning of this paper.

However there is a note of caution to be added when interpreting the analysis layed out in this paper. What we are doing is using machine learning to automatically score the authoritativeness of Wikipedia articles according to community consensus. That last part is the key element to consider, especially in the case where a community is biased. We do not possess such a thing as a 'universal truth reference,' and as such any analysis of Wikipedia will not be able to asertain whether the content is objectively truthful. The best that one can hope to achieve is to judge community consensus regarding an article. In the case of Suh et al.[3], the opinion shared by a group of users involved in edit wars was determined based on how users interacted with each other. Research like this may be incorporated into future work in this field, but it still does not solve the problem of ascertaining truth.

However having said this, we still believe that the the analysis presented in the paper and the results that it has produced are valid and are a significant step forward in aleviating some of the problems faced by Wikipedia and its contributors today. However as with any source of information, the motivations of the source should always be taken into account.

Acknowledgements

I would like to sincerely thank my Supervisor and Mentor for this project, Prof. Pádraig Cunningham and Dr. Derek Greene respectively, for their continued support and advice during this project. The initial idea was for the project was theirs and I could not have done it without their expertise, support and guidence.

I would also like to thank my fellow ODCSSS interns who were located in CASL, they provided help and advice where needed. They were great fun to work and play foosball with.

References

[1] B. Thomas Adler, Luca de Alfaro, A content-driven reputation system for the wikipedia, Proceedings of the 16th international conference on World Wide Web, May 08-12, 2007, Banff, Alberta, Canada

[2] Korfiatis, N. T., Poulos, M., & Bokos, G. Evaluating authoritative sources using social networks: an insight from Wikipedia. Online Information Review (2006), 30(3), 252-262.

[3] Us vs. Them: Understanding Social Dynamics in Wikipedia with Revert Graph Visualizations - Bongwon Suh, Ed H. Chi, Bryan A. Pendleton, Aniket Kittur - IEEE Symposium on Visual Analytics Science and Technology 2007, Oct 30-Nov 1, Sacramento, CA, USA pages 163-170

[4] "Wikipedia:". "Wikipedia:WikiProject History#Showcase". Tuesday 03-Aug-2010. <http://en.wikipedia.org/wiki/Wikiproject_history#Showcase>

[5] József Balogho, Ryan Martin. Edit distance and its computation. The electronic journal of combinatorics 15 (2007), #R20