**International Journal of Emerging Engineering Research and Technology**
**Volume 2, Issue 6, September 2014, PP 150-157**
**ISSN 2349-4395 (Print) & ISSN 2349-4409 (Online)**

# On Measuring Malayalam Wikipedia

**Vasudevan T V**

Asst Professor, Dept of Computer Applications, MES College of Engineering,
Kuttippuram, Kerala, India
*vasudevantv@yahoo.co.in*

**Abstract:** *Wikipedia is a popular, multilingual, free internet encyclopedia. Anyone can edit articles in it. This paper presents an overview of research in the Malayalam edition of Wikipedia. History of Malayalam Wikipedia is explained first. Different research lines related with Wikipedia are explored next. This is followed by an analysis of Malayalam Wikipedia's fundamental components such as Articles, Authors and Edits along with Growth and Quality. General trends are measured comparing with Wikipedias in other languages.*

**Keywords:** *Wikipedia, Malayalam, Quantitative Analysis, Growth, Edits, Quality*

## 1. INTRODUCTION

The Wikipedia is a free and publicly editable online encyclopedia supported by the non-profit Wikimedia Foundation. It was launched on January 15, 2001. Presently it contains 33 million articles in 287 languages. The English Edition of Wikipedia itself contains over 4.6 million articles as compared to 120,000 articles in the next largest English language encyclopedia, *Encyclopedia Britannica Online*. Wikipedia is particularly interesting to research because of the richness and open access of its data. One can analyse wide range of topics ranging from fundamental components to structure and growth of information and author collaboration with little effort of data collection. This paper discusses an overview of research on the Malayalam edition of Wikipedia, which is one of the leading Wikipedias among Indian Language Wikipedias in various quality metrics.

## 2. HISTORY OF MALAYALAM WIKIPEDIA

Wikipedia in Malayalam Language was launched on December 21, 2002. Its URL is http://ml.wikipedia.org. Almost all the early users of Malayalam Wikipedia were non-resident Malayalees. The growth of Malayalam Wikipedia was very slow during the first two years due to the non-availability of a common standard for inputting Malayalam. By 2004, Malayalam Unicode, a Universal Encoding Scheme for representing Malayalam characters came into existence. This led to the development of various malayalam computing tools. Wikipedians started to use these tools and blogging in Malayalam became widespread. Wikipedia reached 100 articles by December 2004. The article count reached 1000 in September 2006. The first major Media coverage about the Malayalam Wikipedia was on September 2, 2007, when Malayalam daily newspaper *Mathrubhumi* covered Malayalam Wikipedia project extensively in its Sunday Supplement. This generated significant interest in the Wikipedia project and large number of users joined the project and started to contribute. The subsequent growth was exponential. The 10000th article was born on June 1, 2009. The mobile version of Malayalam Wikipedia was launched on February 2010. Malayalam Wikipedia presently contains 36774 articles.

## 3. WIKIPEDIA RESEARCH AND MALAYALAM WIKIPEDIA

Wikipedia is one of the most dynamic, popular and large collaborative projects in the Internet. Besides that, It is fully open regarding its editing and accessing policy. Its log files register every single edit performed by Wikipedia authors in any language version. As a result, the databases containing these log entries provide detailed information about the largest open collaborative project we can find today. This is a wonderful opportunity for researchers in different areas such as computer science, sociology, education and linguistics to analyze this project and gain knowledge about its distinct features from many different perspectives.

One can find different research areas related with Wikipedia. Given below are some important areas.

*Quantitative Analysis:* It is a popular research area associated with Wikipedia. Here, we analyse the behaviour of the Wikipedia System using quantitative data. Many of these research works use statistical and data mining techniques for analysing the system. WikiXRay is a python software tool that automates quantitative analysis of all wikipedia language editions.

Some research questions related to this area are:

- What is the total number of articles / authors / words in Wikipedia?
- What is the total / average size of content in Wikipedia?
- What is the number of contributions received in a month / year?
- Compare these quantitative parameters for different language editions of Wikipedia
- Find out the contribution patterns of Wikipedia authors?
- What are the different types of vandalisms affecting Wikipedia?
- Forecast the future trends of Evolution of Wikipedia.

*Analysing Quality of Content*: In this area we try to measure the quality of content in Wikipedia. This is very much important in an open access system such as Wikipedia. Moreover, the problem becomes harder due to the huge size and dynamic nature of the Wikipedia project. Besides, researchers are also trying to develop automated systems that measure the quality of content in Wikipedia.

Given below are some research questions related to this area:

- What is the average quality of articles in Wikipedia?
- Is the average quality improving? Does a typical article improve over its lifespan? How quickly? What trends do we see?
- How can our article assessment system be improved?
- What percentage of articles cites no references at all?
- What policies can we enact to prevent article deterioration?

*Social Networking Analysis:* Large collaborative network systems such as Wikipedia naturally attract social networking researchers. They perform analysis to find out community behavior patterns of different language editions, popularity of content, relationships between content popularity and total number of contributions.

Few research questions related to this area are:

- Which are the most popular articles?

- Which pages are visited together? How close are they in matter of content?
- How many hits per day does the Wikipedia site receive?
- How many readers come from Google? Which pages have high Google Page Rank?
- Identify major topics of interest in Wikipedia.

However, Wikipedia research is mainly focussed on English Wikipedia and other active Wikipedias such as French, Dutch and German. Little or no research has been done on Indian language Wikipedias. This paper discusses about Researching Malayalam Wikipedia, which is one of the most active Indian language Wikipedias.

## 4. ELEMENTS OF MALAYALAM WIKIPEDIA

In this section, after looking at its growth, we discuss about the basic elements of Malayalam Wikipedia followed by a look on its quality.

### 4.1 Growth

Figure 1 : Growth of Malayalam Wikipedia

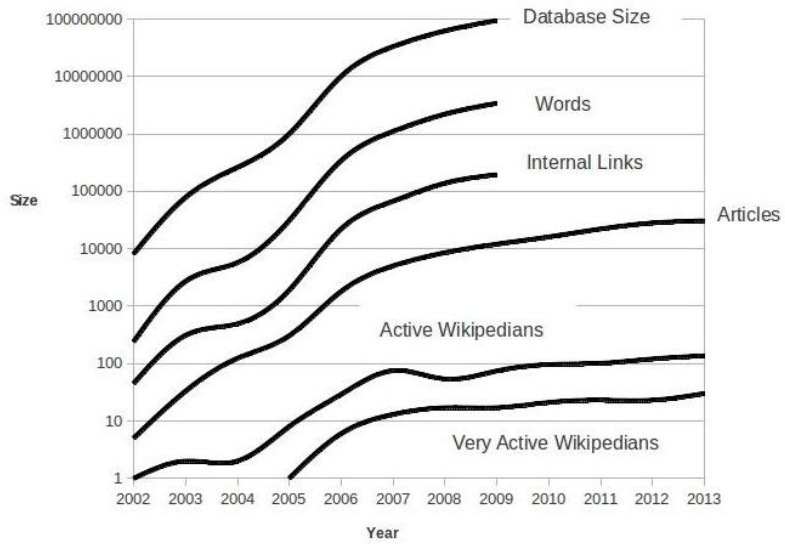


Figure 1 shows six fundamental metrics of Wikipedia's growth. These are:

1. Database size (combined size of all articles including redirects in bytes)

2. Total number of words (excluding redirects and special markup)

3. Total number of internal links (excluding redirects and stubs)

4. Number of articles (at least contain one internal link)

5. Number of active Wikipedians (contributed 5 times or more in a given month)

6. Number of very active Wikipedians (contributed 100 times or more in a given month)

After the initial slow phase from 2002 to 2004, Malayalam Wikipedia began to grow faster. The main reason for this was the standardisation of Malayalam Unicode in 2004. This encouraged many people

to blog in Malayalam. During this phase several Malayalam input tools were also developed, which further enhanced the growth. The biggest growth for almost all parameters was between 2005 and 2006. ( Database size - 75%, Words - 83%, Internal Links - 88% , Articles - 41%, Active Wikipedians - 22%, Very Active Wikipedians - 42%)(Percentage Growth per Month between 2005 and 2006 ). Even though the growth rate has slowed down since then, it has been growing at a steady rate. At regular intervals Wikipedia becomes slow because software and hardware cannot cope with the exploding number of users. Obviously Wikipedia cannot grow infinitely but at the moment there is no evidence that it will not continue growing.

**Table1.** *Malayalam Wikipedia at a glance July 2014*

| | |
|---|---|
| Page Views per Month | 4,687,856 |
| Article Count | 36,843 |
| New Articles per Day | 10 |
| Edits per Month | 7,688 |
| Active Editors | 87 |
| Very Active Editors | 15 |
| New Editors | 14 |
| Speakers | 37,000,000 |
| Editors per Million Speakers | 2 |

**Table2.** *Top 10 Wikipedias of The World July 2014*

| ***Rank*** | ***Language*** | ***Article Count*** | ***Very Active Editors*** | ***Active Editors*** | ***Editors per Million Speakers*** |
|---|---|---|---|---|---|
| 1 | English | 4,656,030 | 3,037 | 31,819 | 21 |
| 2 | Swedish | 1,798,095 | 97 | 648 | 65 |
| 3 | Dutch | 1,784,835 | 207 | 1,260 | 47 |
| 4 | German | 1,718,247 | 851 | 5,944 | 32 |
| 5 | French | 1,532,774 | 721 | 4,409 | 22 |
| 6 | Italian | 1,141,981 | 363 | 2,675 | 38 |
| 7 | Russian | 1,139,353 | 529 | 3,286 | 12 |
| 8 | Vietnamese | 1,122,516 | 51 | 307 | 4 |
| 9 | Spanish | 1,117,161 | 484 | 4,142 | 8 |
| 10 | Waray-Waray | 1,073,774 | 1 | 16 | 5 |

**Table3.** *Top 10 Wikipedias of India July 2014*

| Rank | World Rank | Language | Article Count | Very Active Editors | Active Editors | Editors per Million Speakers |
|---|---|---|---|---|---|---|
| 1 | 50 | Hindi | 102,877 | 13 | 57 | 0.1 |
| 2 | 62 | Tamil | 62,181 | 15 | 75 | 1 |
| 3 | 63 | Telugu | 58,792 | 16 | 57 | 0.7 |
| 4 | 69 | Urdu | 53,010 | 2 | 23 | 0.4 |
| 5 | 76 | Marathi | 41,300 | 5 | 28 | 0.3 |
| **6** | **78** | **Malayalam** | **36,843** | **15** | **87** | **2** |
| 7 | 84 | Bengali | 31,423 | 15 | 66 | 0.3 |
| 8 | 96 | Gujarati | 25,570 | 0 | 5 | 0.1 |
| 9 | 97 | Bishnupriya Manipuri | 25,230 | 0 | 0 | 0 |
| 10 | 104 | Kannada | 17,935 | 4 | 22 | 0.5 |

### 4.2 Articles

Every article in Malayalam Wikipedia is referenced via a unique name. You access it with a URL like http://ml.wikipedia.org/wiki/name-of-article, where the subdomain 'ml' corresponds to Malayalam Language Edition of Wikipedia. Synonyms can be redirected to another article. Articles can be directly edited without any knowledge of HTML using a special Wiki syntax which can be learned quickly. Each article deals with only one concept so that article titles form a controlled vocabulary. Malayalam Wikipedia also contains heavily linked articles for years, day of a month, decades, etc. which constitute an almanac.

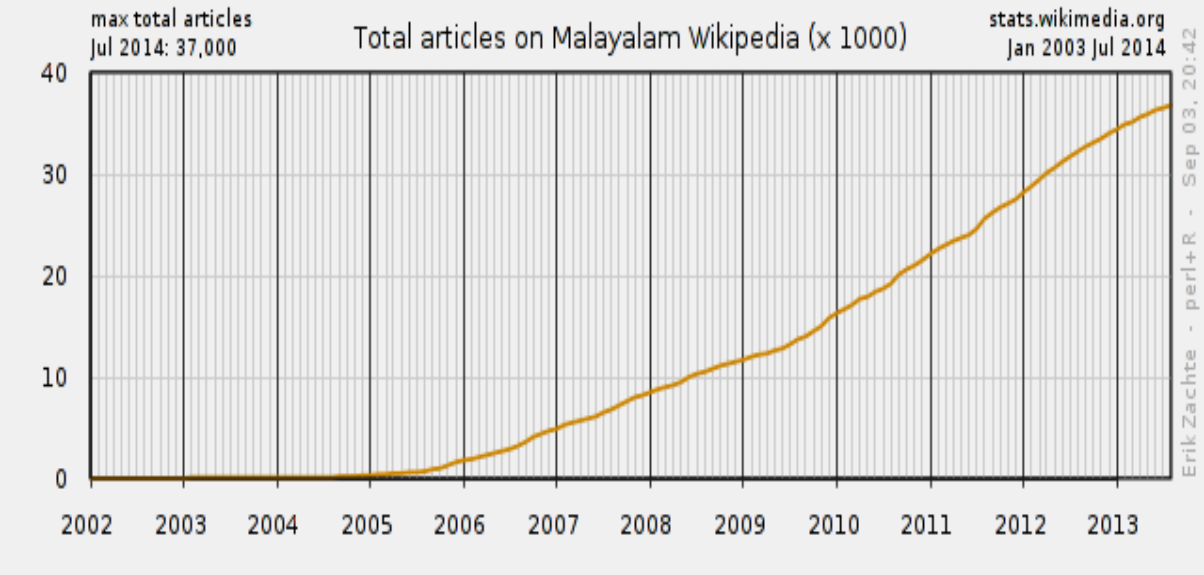


**Fig2**

A large percentage of the articles may be short stubs, which cannot provide contents of encyclopedic nature. They can be merged to form good articles. In Malayalam Wikipedia, the average size of an article is 2897 bytes compared to 3655 bytes in English Wikipedia. 38% of the articles in Malayalam

Wikipedia have at least 2 Kb readable text, while in English Wikipedia this is 45%. In Malayalam Wikipedia 89% of articles have at least 0.5 Kb readable text as against 91% in English Wikipedia. One can see that, in these parameters, Malayalam Wikipedia is very much close to English Wikipedia.

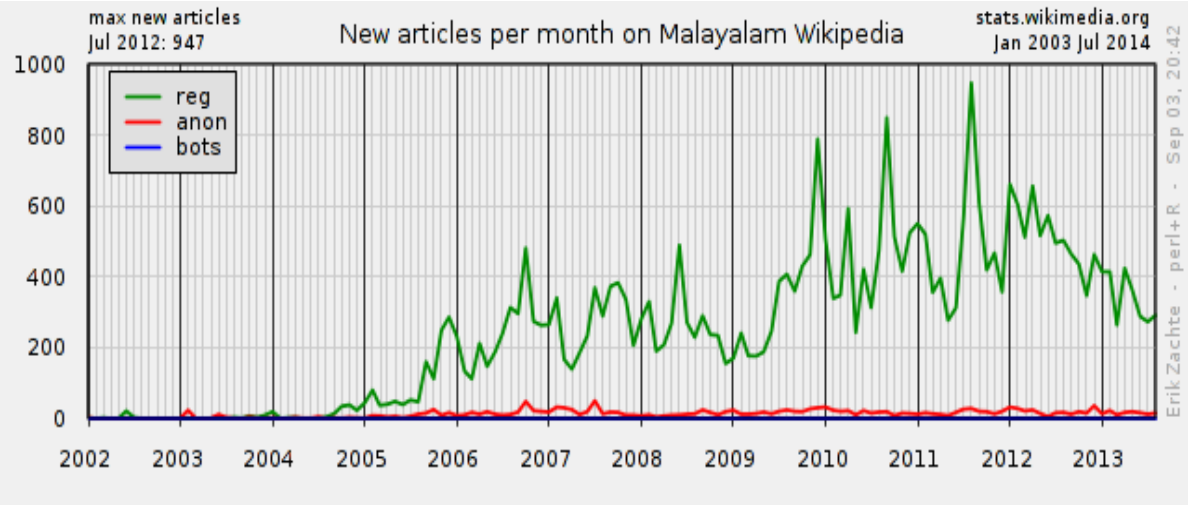


**Fig3**

From 2006 to 2009, on an average, 300 new articles are created per month in Malayalam Wikipedia. This count increased to around 600 from 2010 to 2012. In July 2012, 947 articles were contributed to Malayalam Wikipedia, which is the highest ever contribution.

### 4.3 Authors

As everyone is invited to contribute, Wikipedia articles can have a large number of authors. An author who performs edits in a Wikipedia article can be a registered user, an anonymous user or a bot. A bot is an automated or a semi-automated program that performs edits in Wikipedia to carry out repetitive and mundane tasks. There are also administrators or sysops who can block user accounts or IP addresses from editing, protect pages from editing etc. to prevent vandalism.

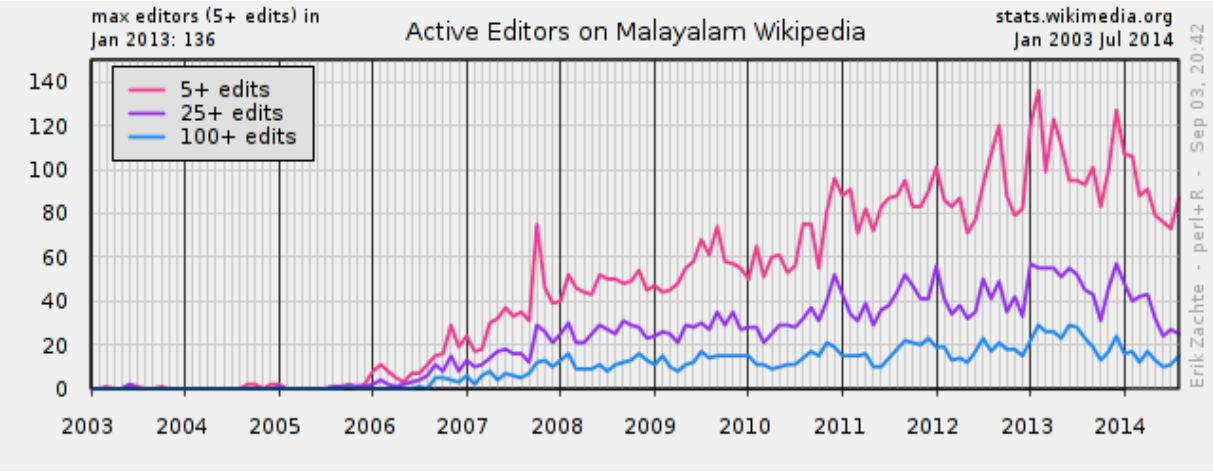


**Fig4.**

Malayalam Wikipedia has the highest number of active editors among Indian Language Wikipedias (Refer Table 3). We can also see that the participation rate of Malayalam speakers in editing is highest among Indian Language Wikipedias. There are 5693 registered editors (minimum 1 edit) in Malayalam Wikipedia, in July 2014. On an average 14 new editors (minimum 10 edits) enter Malayalam Wikipedia every month, since January 2013.

### 4.4 Edits

When an editor changes an article, his edit is recorded and gets listed in the article's version history where one can highlight differences between selected versions. Users can add articles to their watch

list to get informed on changes. All changes are listed at the recent changes, where one can observe new contributors and suspect edits. Sometimes two authors may revert each others edits leading to a phenomenon called ‘edit war’.

There has been 1,101,632 article edits in Malayalam Wikipedia upto July 2014. Among these 31,958 article edits (3%) were made by anonymous users.

**Table4.** *Distribution of Article Edits Over Registered Editors in Malayalam Wikipedia Upto July 2014*

| **Minimum Edits** | **Registered Editors** |
|---|---|
| 1 | 5693 |
| 10 | 1226 |
| 100 | 299 |
| 1000 | 102 |
| 10000 | 17 |
| 31623 | 3 |

### 4.5 Quality

Compared to other encyclopedias like *Encyclopedia Britannica* which are edited by experts, Wikipedia can be edited even by a novice. Therefore, an important area of concern associated with Wikipedia is the quality of articles. However, due to its open nature, when more people read an article, more errors get amended. But one can hardly be sure how many qualified people have read and edited articles and how many errors remain.

Several studies have been done to determine the reliability of Wikipedia. A notable study by the journal *Nature* conducted in 2005, found that Wikipedia scientific articles came close to the level of accuracy in *Encyclopedia Britannica* and had a similar rate of "serious errors".

In Wikipedia, there is a procedure to select articles which are then reviewed and improved by editors until they get “featured article” status. These articles are verified for accuracy, neutrality, completeness, and style.

Malayalam Wikipedia contains 136 featured articles as compared to 4379 in English Wikipedia. The Featured Article Percentage (No of Featured Articles / No of Articles * 100) of Malayalam Wikipedia is 0.37, whereas this is 0.095 in English Wikipedia. On an average around 30 articles are selected as Featured Articles per month in English Wikipedia, while this count is approximately 1 in Malayalam Wikipedia.

## 5. CONCLUSION

This paper presented an overview of Malayalam Wikipedia Research**.** After discussing the relevance of Wikipedia research in general, we looked at the basic trends in the growth of Malayalam Wikipedia. Then we compared various parameters of Malayalam Wikipedia with other language Wikipedias to know its present position among other Wikipedias. We have understood that it is one of the leading Wikipedias among Indian Languages. We have also seen that it is even closer to English Wikipedia on some parameters. There are several possibilities for further investigation such as quality, content etc.

## REFERENCES

[1] Jakob Voss, "Measuring Wikipedia", Proceedings of the ISSI 2005, July 2005.

[2] Rodrigo B. Almeida, Barzan Mozafari and Junghoo Cho, "On the Evolution of Wikipedia" , International Conference of Weblogs and Social Media, Mar 26-28, 2007

[3] Felipe Ortega, Jesus M. Gonzalez-Barahona and Gregorio Robles, "THE TOP-TEN WIKIPEDIAS - A Quantitative Analysis Using WikiXRay", ICSOFT 2007

[4] Jos´ Felipe Ortega Soto, "Wikipedia: A quantitative analysis " , Doctoral Thesis, Madrid, 2009

[5] Anselem Spoerri, "What is Popular on Wikipedia and Why " , First Monday (Peer-Reviewed Journal on The Internet), Vol 12, No 4, April 2007

[6] Dorde Stakic, " Wiki Technology – Origin, Development And Importance ", INFOTHECA – Journal of Informatics and Librarianship,No 1-2, vol X, June 2009

[7] Dennis Wilkinson and Bernardo Huberman, "Cooperation and Quality in Wikipedia", International Symposium on Wikis, 2007

[8] Stats.wikimedia.org

[9] http://ml.wikipedia.org/wiki/

[10] http://en.wikipedia.org/wiki/Malayalam_Wikipedia