Enhancing the Trustworthiness, and Comprehensiveness of Wikipedia Content

Susanna Mkrtchyan
Institute for Informatics and Automation Problems of NAS RA
Yerevan, Armenia
e-mail: susanna.mkrtchyan@wikimedia.am

Abstract—**The approach proposed in this article is aimed to gradually design and build a system of gadgets and instructions for effectively coordinating and synergizing online and offline efforts of users, to guide their contributions towards trustworthy, comprehensive, deeper, and more complete development of all language versions of Wikipedia.**

Keywords— **Wikipedia, content, trustworthy, small languages' maintenance.**

I. THE GAP BETWEEN LANGUAGE VERSIONS

Wikipedia -Imagine a world in which everyone on the planet is given free access to the sum of all human knowledge. —Jimmy Wales

It also means you have access to the entire sum of knowledge created by humanity, in the language, which you own.

Nevertheless, there is a huge gap between the content of language versions of Wikipedia in quantity, even more on quality [1].

It is a big challenge for small languages to stay used in educational, professional and scientific areas [2--3].

Looking at Wikipedia's statistical table of languages (see Table 1) one can see a huge gap in quantity, and even more in quality [3].

Only several languages can be compared to the English Wikipedia, but even they are multi-times smaller in volume.

The volumes of the content of languages lower in rank are even smaller by several tens or hundreds of times [1].

There is a lot of criticism of Wikipedia: poor, not trustworthy. Teachers and specialists, in general, forbid their students to use Wikipedia. Meanwhile, students continue to use Wikipedia. A question arises, why teachers and professors do not correct or develop Wikipedia? But…

It is not an easy task to catch up on where, what, and how one can do, to correct or just add information that is missing.

II. IMAGINE WIKIPEDIA A SET OF THEMATIC WIKIPEDIAS

The approach proposed in this article aims to fill gaps in the content of any language version of Wikipedia, as well as among versions in different languages. The key point of the suggested approach is imagining Wikipedia as a set of Wikipedias on thematic areas, which allows to group users via their knowledge and interests, so as to collaborate more effectively.

It is necessary to gather users/professionals of the same area together around the exact portal, to facilitate their efforts to find exact topics needed to create or update, to provide flexibility for users to coordinate their efforts on enhancing the content of Wikipedia of a certain area by coordinating, organizing, and monitoring the content of that field in Wikipedia.

All required information is maintained in Wikipedia. Which category of articles does the participant edit, which articles are newly written and not checked, or which articles are missing.

III. PORTALS AS A PLATFORM, VILLAGE PUMP FOR THEMATIC WIKIPEDIA

The approach suggested here is gathering all this information about the current situation of the content in a certain area, and its editors and their coordinated actions to put one page, research monthly feedback for generating balanced actions.

Wikipedia already has such pages - thematic portals [4], which can be improved and executed more actively and effectively.

Below is the suggested structure of the Portal page:

- The list of currently missing article.
- The list of feature articles and candidates.
- The list of good articles and candidates.
- The list of upcoming events related to the topic, it can be the date of online meeting or editatons, campaign, etc.
- The list of articles to be reviewed.
- The list of articles to be developed.
- The list of existing articles and their quality evaluation.
- Organization and discussion of articles' review.

Table 1:

language	wiki	article	page	edits	admin	user	acti	characte	depth
English	en	6 633 727	57 811 306	1 141 399 445	910	45 239 307	130 416	898 426	1175.08
Deutsch	de	2 784 507	7 686 209	230 550 086	182	4 113 710	18 975	127 403	92.95
svenska	sv	2 559 809	6 159 287	52 125 341	67	859 190	2299	0	16.73
français	fr	2 507 367	12 481 487	202 081 582	154	4 615 324	19 265	70 293	256.2
Nederlands	nl	2 119 052	4 522 145	63 872 232	35	1 265 811	3956	20	18.16
русский	ru	1 902 773	7 630 161	129 081 029	75	3 355 889	11 040	242 774	153.27
español	es	1 848 552	7 867 383	149 603 744	60	6 797 853	14 583	0	201.59
italiano	it	1 803 332	7 674 555	132 331 620	119	2 361 469	8317	133 474	182.77
polski	pl	1 561 204	3 598 360	69 716 983	99	1 226 914	4496	261	32.99
日本語	ja	1 367 388	4 026 466	94 234 090	41	2 050 129	14 687	7889	88.5
中文	zh	1 342 253	7 330 786	76 159 447	67	3 327 175	8237	62 815	206.8
Tiếng Việt	vi	1 281 935	19 412 992	69 751 487	19	902 730	2007	25 590	718.75
українська	uk	1 251 743	4 335 491	38 910 035	46	678 512	3957	111 755	54.47
العربية	ar	1 203 379	8 047 317	61 799 589	20	2 396 187	4608	51 515	248.39
português	pt	1 103 042	5 485 134	65 357 363	52	2 891 048	9179	59 264	188.06

- Language connectivity of the portal pages will also make it possible to get acquainted with the work done in more developed portals and to copy the good practices, as well as to discuss and ask for help.
- The list of subscribed members to the certain portal.
- The list of articles under development and/or review for each participant.
- The rules for becoming a member of the portal and who is eligible to vote on article quality.

After the creation or modification of the Portal page in such a structure. We will need gadgets to keep all fields updated, so one can subscribe and get assignments/offers to do.

The only thing left to do is to push, suggest or regularly remind the editors of the relevant category which article is waiting for revision, which ones need improvement. Just as YouTube offers music according to your taste, Netflix offers movies, you may be have interested in.

Based on the information on Wikipedia statistics and the feedback of outside users and professionals, the system can improve itself from time to time.

IV. SEMI-AUTOMATIZATION

The creation and renewal of portals, as well as the attraction of contributors into the coordinated work can be semi-automated over time by entering into a dialogue with newcomers and active editors in the given field:

- suggestion to sign as a member of a certain thematic community to the portal page;
- attracting to the portal page and its update;
- suggestion to create an article;
- review a new or recently enlarged article;
- discussing everything related to a certain field of Wikipedia.

V.   CONCLUSION

There are hundreds of communities that are devoted to developing their home Wikipedias, believing that it will help to keep using their native languages.

But statistics show that only several language visits are increasingly appropriate for their development.

Users prefer comprehensive ones, where they can find answers to all their requests (searches).

So far, the smaller languages can't be compared to the largest ones.

The approach suggested in this article is a way to work on filling the gap of the content of small languages, so that they can be more developed, even comparable to larger ones. It can help them to reach more visits and, therefore, will keep the language alive in all areas.

REFERENCES

[1]   [Online]. Available:
       https://cacm.acm.org/magazines/2021/4/251343-building-a-multilingual-wikipedia/abstract
[2]   [Online]. Available:
       https://vc.bridgew.edu/cgi/viewcontent.cgi?article=1593&context=honors_proj
[3]   [Online]. Available:
       https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24553
[4]   [Online]. Available:
       https://en.wikipedia.org/wiki/Wikipedia:Contents/Portals