

Expert interview

Case: Wikipedia
Interviewee (IE): Amir Aharoni
Interviewer (IR): Anja Ebersbach
Date: 2021-07-16, 11:00am
Location: Skype

	[Welcoming]
1	IR: Ok. Let's go on with our interview. Yesterday I checked our last interview in preparation for today and we stopped when you told me about the quite complicated process of creating a new Wikipedia, a new language version. You told me what's necessary to simplify that. Is it on technical base? #00:00:58#
2	IE: I will just send to you these links. I have some written texts and video recordings about this. So, before I forget it, just we just do this. #00:01:39#
3	IR: Yes, you told me about the Celtic knot. #00:01:39#
4	IE: Yes, I found it. This should be the beginning. Watch the video and there is also a Phabricator task and from the Phabricator task there are also a few more links that you may find interesting. The task description is pretty long and there are also links to another Wikimania talk by somebody else. I think he is German, but he spent many years living in African countries, in Namibia and in some other places and I am pretty sure he knows German. Dieter Gallert is his name, maybe you want to speak with him. There are links to a talk he made in Wikimania, where he speaks about the same problem. He does not offer a detailed solution, but he describes the problem quite well. He is an experienced Wikimedian and he wants to help other people who speak other languages to set up their languages on Wikipedia and it's difficult for them and it's difficult for him. And what I propose is to address this difficulty and make it easier. #00:02:30#
5	IR: That is a good contact, I am going to ask him. #00:04:43#
6	IE: He is not the only one who complains about this, there are lots of people who complain about this. He just made a whole tech talk on Wikimania about this. He also speaks about other topics in that talk but this is one of the major topics of his presentation, so watch that. He presents it quite well. You can probably find him and speak to him directly. #00:04:48#
7	IR: Is he more active in the English Wikipedia or in the German one? #00:05:14#
8	IE: He is in both. Let me check very quickly. It's very easy to find. The username is DGallert. Mostly English I see. He also made some contributions in German, but it is mostly in English. #00:05:25#
9	IR: Ok. Good. Perfect. Let us go on. #00:06:08#
10	IE: Just one second.... Yes, I am back with you. #00:06:12#
11	IR: Does Wikipedia have a multilingual search engine? #00:06:36#

- 12 IE: Kind of yes. For that I can give you another contact. He is the engineer who maintains it, there are several, but he is the one who is a really serious expert in search engines in general and in multilingual support in particular. His name is Trey Jones. He is a real scientist. I am a linguist, too, but he also knows computational linguistics properly. I never learned computational linguistics and he is very good at writing and explaining these things. #00:06:43#
- 13 IR: He has written about that already? #00:07:54#
- 14 IE: Yes, a lot and you can talk with him about this. The engine definitely exists, the problem with that engine is that the user interface is far, far from perfect. Not as convenient as it could be. In theory it is possible to make a search engine where you will go, for example, to the German Wikipedia and you will search in the search box in English and it would show you results for English. The technology exists and does not require any significant additional coding, but it's not properly enabled everywhere. I don't know why. I think it needs not so much engineering, it doesn't need any complicated software development. It just needs some design and product management. Because the technology exists. Because all the languages are indexed and there is even morphology support for some of them. Not for all of them, because we support more than 300 languages and we do not actually support proper morphology and search in all languages. But we support proper morphology in some languages and Trey Jones is the perfect person to speak about this because he is the expert. He is working on integrating these things. #00:08:55#
- 15 IR: What you want to say is: the search is there and nobody knows? #00:10:37#
- 16 IE: It kind of exists, but it gives weird results. I spoke about this several times. I can even send you some example of unresolved bugs. There are some weird features. Some Ukrainian speakers complained about this to me... lots of people know that they can talk to me about bugs, I won't be able to resolve the bug, but I can help them report the bug to the right person. So, they spoke to me and they gave some examples: if they search for Ukrainian words in the search box on the English Wikipedia, they find results from the Russian Wikipedia even though there isn't an article about this in the Ukrainian Wikipedia. And the name of the article is exactly the same. They use the same alphabet and many words are spelled exactly the same or pretty similar. For this particular word it was the exactly same word. There is an article with the same name and written with exact same letters but the search engine finds the Russian Wikipedia entry but does not find the Ukrainian Wikipedia entry and also does not find the Bulgarian Wikipedia entry. Even Bulgarian was also written with the same alphabet and has an article about the same title. I spoke about that with engineers. Currently it works as designed, right? So I guess the design is not good because the fact is that people complain about this and they cannot find what they are looking for. #00:10:44#
- 17 IR: Is it a Wikimedia project then... how is it developed? Is it a community thing? #00:12:54#
- 18 IE: It is maintained by the team in the Foundation, there is a team of engineers. Recently, few months ago a new product manager was hired. Because before several years they had several engineers but they didn't have a product manager. Or designer. So that's why the engineers are not enough. You need some engineers to write the code but you also need the product manager and designers, that is not

same thing. So recently they hired a product manager, so I hope this product manager can address this problem. It's really an issue of product management. The technology it's all there. It all exists already; they do not need any new developments. #00:13:04#

19 IR: So, it's coming? #00:13:49#

20 IE: Its totally coming. It's not a question for me when it's coming, I am not the product manager. Thats is a question for the product manager. As a general comment: I think there was some data collection research about this how often this happens. But I see people do this all the time. They go in Wikipedia in one language and they search in the search box in other language. They go for example in Wikipedia in Hebrew and they search in English. Or they go in Wikipedia on English and search in Russian. So, it's very common. People do this all the time. And there is technology to show the results from another Wikipedia, from Wikipedia in another language, it exists. It is even enabled on the projects, it just has some design problems. That's why people try to search in Ukrainian and they find Russian. Even it should be possible to find in Ukrainian. It's kind of enabled but it has a lot of bugs and design problems. #00:14:59#

21 IR: You can immediately see that it is a different Wikipedia or you just have to go to...? #00:16:50#

22 IE: Yes, give me just a moment. I remember the word. It should be familiar word for you too, it is transistor. The electronic component. Let me just try to share the screen. Can you allow me? #00:17:20#

23 IR: Yes. #00:17:48#

24 IE: Because of that I probably lost my permission to share the screen. #00:17:49#

25 IR: Now? #00:17:58#

26 IE: Yes. You should see English Wikipedia. Now I go to the search box. I switch my keyboard to Russian and I write tranzistor. There is no such an article on English which make sense because it's English Wikipedia and doesn't find it. But. If I now press „enter“, let's see what happens.

27 IE: You see, it's showing results from the Russian Wikipedia. It doesn't find something in English Wikipedia, which is sensible. But it shows results from Russian Wikipedia. You see on the side bar there are also results from the English Wictionary and it has the Russian word in English Wictionary, this is also quite sensible but there is such an article in the Ukrainian Wikipedia. So, if I go here, I click the link to the Russian Wikipedia. This is Russian, but if I change URL instead to UK, which is Ukrainian. And you see the rest of URL is the same. There is an article, it's about the same thing. #00:18:04#

28 IR: But it's not shown as a result? #00:20:00#

29 IE: No, it's not shown as a result in the English Wikipedia. I will expect that the English Wikipedia shows both of them. There are also couple of other languages, there is Bulgarian, I will definitely expect to see Bulgarian here... let me try BG. Yes, the same thing, you see: the same title, that is Bulgarian.... If I try KK, which is Kazakhstan, they use similar alphabet. Yes, also Kazakh. So, this is a problem. It

	only shows Russian. Of course, the Russian is the biggest language with the most speakers but Bulgarians do not know Russian. It's not very good. #00:20:03#
30	IR: But then it should work for many things, shouldn't it? For names, for example. #00:20:59#
31	IE: Yes, it's just one example. Somebody complained for this word, but this will probably happen for a lot of other words. ##00:21:05##
32	IR: If I search for Angela Merkel it will show me probably only the German Wikipedia, but she is also represented in so many Wikipedias. #00:21:09#
33	IE: If you go to the English Wikipedia and you type Angela Merkel in the Latin alphabet you will see the English Wikipedia article. But if you type in Cyrillic alphabet... let's see what happens... ok, so, Russian Wikipedia, Angela Merkel, that's fine and that's it. Let's go to URL and let's see what's happened when I change in UK ... I think the title is not exactly the same in Ukrainian, but it's pretty similar. In Russian Wikipedia they write last name in the beginning. That is the first title that I search in English Wikipedia and it should find Ukrainian but it doesn't find Ukrainian. #00:21:35#
34	IR: But if I type Angela Merkel in the English Wikipedia is normally should show the German article or the French. #00:23:24#
35	IE: No, it will immediately show you the English article and it will show interlanguage links on the side bar. If you go to the English Wikipedia and type the name in Latin alphabet and there is such an article, then it's quite obviously that will show that article... that's fine. And that will show interlanguage links in the side bar. I think this is fine. #00:23:37#
36	IR: But this is the parallel situation to your Russian case... if you search for a name in the Russian Wikipedia and there is a Russian article... it's just like typing Angela Merkel in the English Wikipedia and there is an English article.... Isn't it the same situation... because they say it's the biggest Wikipedia and probably you want to read the Russian version? #00:24:11#
37	IE: Hmmm... well... that's.... really... I do not have good answer for that. That is really a question for the product manager. A product manager who will do proper design and research what does really work for people. #00:24:41#
38	IR: Are there numbers for how many Wikipedias, how many versions are closed or which one are closed? I try to find.... #00:25:08#
39	IE: That is very easy. Just go to Meta and search for list of Wikipedias and there is a list there. #00:25:23#
40	IR: Also the ones which do not exists anymore? #00:25:35#
41	IE: Yes, yes. #00:25:40#
42	IR: Is there a date, when they were closed? #00:25:46#
43	IE: Yes, I think so. #00:25:51#
44	IR: Good. Do you know any other organizations outside of the Wikipedia

	ecosystem which are interested in language topics? I think of Goethe Institute or the equivalent for that in other countries. Do you know? #00:25:55#
45	IE: It definitely exists. For some reason I never spoke to anybody about this as much detailed as I spoke about this with you. #00:26:20#
46	IR: Really? #00:26:33#
47	IE: Yes, like I spoke with you about this for several hours. And I never spoke about this in so much details to anybody, but definitely there are organizations that are interested in this. Google is interested in this. I spoke with Google about this, I have one interview with them. The problem with Google is that it is an extremely non-transparent organization – company. They care a lot about automating everything and doing everything at a huge scale. And they don't care if along the way something is wrong. They can tolerate 20 percent of wrong information as long 80 percent of information is somehow useful. And they don't care. And maybe even more than 20 percent and they don't care. They do not like to make small adjustments, or fixes, or talking directly to people about problems. They want to do everything big. And the Wikimedia team is the exact opposite. In Wikimedia almost everybody likes to do things really carefully and meticulously and manually...and check every fact and fix every spelling mistake and so on. And Google is really not like that. #00:26:52#
48	IR: So, it doesn't fit. #00:28:40#
49	IE: Yes. #00:28:42#
50	IR: Ok. So...one of the last questions: If you think about Wikipedia and multilingualism, what are the great strengths and what are the weaknesses of this system and what is something that you would change immediately if you had the money and the people and the go from the WMF? #00:28:56#
51	IE: Well, two things. One of the more technical product level and I am pretty sure we talked about this few weeks ago and templates. The fact that templates can't be shared across different wikis. This is the number one, like top priority. The fact that this cannot be shared that means that different languages cannot use the same software. This is extremely inconvenient, it's extremely unjust, it's like... you should give same software to everybody - and how people use software, it's a different question - we're supposed to give everybody same tools and in practice people do not use the same tools, so this is very bad. That is on the more technical product level. On the more social level, social-cultural level the whole Wikimedia movement says that its vision is imagine the world in which every single human being can share in the sum of all knowledge, every single human being. That is the motto and they love this motto. I completely support this vision, it's a wonderful vision. Not everybody cares about that very strongly, I do care about that very strongly and I remember it by heart. A few years ago, I attended some training about product management and they spoke about the importance how every company supposed to have a vision. Like the vision of Google is to organize the world's information. And the vision of Coca Cola is to give everybody happy moments or something like that's Coca Cola. So, they ask in every company that you work: do you know what is the vision? I was the only person in the room that knows what is the vision of the company. #00:29:22#

52 IR: Really? #00:31:45#

53 IE: Yes, like apparently there are product managers that don't know what the vision of the company that they work for is. I was the only one who knows the vision and I care about this vision. The problem is: this vision it is orthogonal to reality. This vision is relatively easy to achieve for languages and cultures. There is a lot of overlap but it's not perfect 100 percent overlap between language and culture. But let's say partial overlap. It is easy to achieve for languages and cultures in which there is a tradition of at least several decades of having public education, and libraries and academic publishing and things like that, in that language. So, you can easily do that in English, which have several centuries of public education and academic publishing and printed encyclopedias and online encyclopedias and newspapers. It's easy to do this in German, and in French, and in Russian, and in Polish, and in Chinese. It's really easy to do this in Hebrew. Hebrew has thousands of years a literal tradition and more than a century of academic publishing in Hebrew, so it's very easy to do this in Hebrew as well and in several dozens of other languages. It's more difficult to do this in Swahili language, even though a lot of people speak Swahili. Swahili is enormous language, it's difficult to count how many people exactly speak, but it's at least 50 million, probably more. It's an enormous language by numbers of speakers. But relatively few of these speakers actually use this language for reading educational literature. There is some educational literature but not nearly as much as in German. And most of the people who even want to read educational literature, they probably know English. If they do not find something in Swahili they will just read it in English and a lot of them won't even try to look for it in Swahili. They will just go directly to English. In countries like Tanzania, and Kenia, and Congo. And that is a problem. Because for people who know only Swahili and don't know English or French, they do not have anything. That is the major cultural-social problem. So, some people in the Wikimedia movement and around, they just said that is how it is, there is limit how much we can change the human culture and economy and social condition of humanity. I try to be more adventurous. I try to be more radical, as much as I can, and I try to say: ok, we already achieved this goal to let in almost everyone who knows English and German and Hebrew and Russian to share the sum of all knowledge, like pretty much everybody who speaks these languages can find what they need on Wikipedia. You have some edges. Very common example is that in India and Philippines and some African countries a lot of people speak English, maybe they also speak a local language but they also speak English and they cannot find what they need in English Wikipedia because English Wikipedia's rules like what article should be there and should not be there, they are too strict so they cannot write articles about African actors in English Wikipedia. Because they cannot find proper references, or write them according to the style of the English Wikipedia, because American and British notability requirements are dominant in the English Wikipedia and they clash with the needs of people who speak English in Nigeria, and Ghana, and South Africa, and Kenia and India. So, there is this edge, right? #00:31:48#

54 IR: What are your ideas to help the African Wikipedias? #00:36:54#

55 IE: There are several things that we could do, like some people say not to give up. Some people say they are not looking for information in Swahili, in Urubu and in

these languages, they know English anyway...this is factually wrong. Not everybody there knows English. They say: the people who want education they will learn English first anyway. This is wrong, it's almost racist. It's awful. 200 to 300 years ago, people started publishing books in German and you could say why we just do it on Latin, everybody knows Latin, everybody who wants education knows Latin anyway. And if you say that today it is nonsense. Probably there were some people two hundred years ago who said such a thing. Now people say the same thing about Swahili and Hebrew and if you want education you need to learn English anyway. This is not my attitude. In terms of practical things, the Wikimedia Foundation, the Wikimedia movement, the Wikimedia chapters can do about this several things: I am not afraid of paying some people to write articles. I do think that you have to very carefully choose what to write about and who to pay for this, but generally I think it's reasonable. Lot of people say, yeah, it is fully ok, but some people maybe dislike that, some people may be unhappy that they are volunteers and they are doing it for free and some other people are paid. So, the volunteers will be less encouraged to write for free. I haven't heard anybody who says that about themselves. I haven't heard anybody who says: if I see you start paying people in Africa to write articles in Swahili, I will stop writing in French because of this. I never heard anybody say this about themselves. So, I think it is an acceptable idea, you do have to be thoughtful about who do you pay and what they do write about and so on. But generally, I think it's a totally possible thing to do. Think about the fact that the first articles in English and German and Russian Wikipedia and also in other languages, they were actually copied from all Encyclopedias in these languages. A lot of the first, like, few thousand articles in the Russian Wikipedia, they were copied from the Brockhaus Encyclopedia. There is an old out of copyright edition, Russian edition, of Brockhaus Encyclopedia and they just copied it. Somebody paid for this. It was not paid directly for writing Wikipedia. But somebody paid for this 100 years ago. And now the Russian language has a privilege for having some place for copying encyclopedia articles from. And Swahili doesn't have this. So, I think it's totally conceivable that the Wikimedia Foundation will do this or will be that some other organization will do this, or NGO, I don't know, SIL or UNESCO or whatever ... translators without borders, I don't know. That sounds like totally sensible thing to do, to pay people to write articles.p
#00:37:00#

56 IR: One question about this. Would you pay people to write an article or to translate an article? #00:40:59#

57 IE: Both. This thing, Brockhaus Encyclopedia, it was published little over a century ago and you probably heard about the German Brockhaus Encyclopedia, it is a very old German brand. In Russia some people wanted to publish a big Encyclopedia in the Russian language. They decided that the easy way to do this would be to translate from German. Lots of people back then, Russians and Germans, they felt that it was a good Encyclopedia and they can translate that. So, they started to translate articles, they translated articles that were easy for them but on some point, they realized that there are some topics that are useful for people in Russia and are just not covered at all in the Brockhaus German Encyclopedia, so they have to write it from the scratch and for some of them, the style of the German article was not convenient, so they wrote something by themselves. The same thing, today, I use to translate from English, German, French, and Russian to other languages and some

articles you have to write from scratch. And I think it's the best to let those people who speak these languages to decide what they want to do. In all these languages you can find educated people: teachers, linguists, journalists or writers or judges. You can even organize like pay some committee or to have 10, 15, 20 people knowable, respected people in their communities, large communities of people that speak these languages and let them compile a list of articles. You can get from several sources like articles that would be popular in that country. Articles like the 2.000 most important people of the history of Nigeria. To make sure there are articles about them in the Urubu language, there are several languages in Nigeria, Urubu is one of them. If they can be translated from English and the translation will be good and useful for Urubu speakers that's fine. If they have to be written from scratch, then write them from the scratch. That's also fine. It's case by case. That's another example how we in Wikimedia movement want to be meticulous and detail oriented. Google would not do this. Google wants to do something huge and they don't care if half of it is garbage. We thought we can do this; I think we have a proof we can do this. We have been doing this for 20 years and it has worked pretty well. In other languages. Doing it with less advantaged languages will have some challenges but we achieved so much till now and we can do more. #00:41:07#

58 IR: Are you alone with this opinion? #00:44:29#

59 IE: I am not alone. I am definitely not alone, I remember some particular cases, and there was Wikimania in South Africa three years ago. To every Wikimania people come from some African countries, but because it was in South Africa in 2018 there were even more people from African countries than usual. Which is great, of course. And there was one African woman there who just made a passionate speech there saying: it's just not right that you expect that people in Africa to volunteer and write for Wikipedia. They do not have the tradition, they don't have the time, they barely have something to eat...and you want them to work for free and write article. She is "fight" and I don't want to argue with that, she is fine. And a few years ago, there was a longish thread on one of the mailing lists. Somebody said: why don't we pay some people in some languages to write and translate articles? Does anybody disagree? And nobody really disagreed. I said I agree and a lot of people say, yeah, it's totally ok. I don't know why we are not doing this. I am definitely not alone. #00:44:36#

60 IR: So, thank you very much. I think we have finished.

[Goodbye]