

Final Report – Project in the course - Digital Humanities For CS Students

1. Research Question:

Background: In general, the Hebrew Wikipedia is much less extensive and considered less qualitative than the English Wikipedia. On the other hand, regarding more esoteric issues or topics that are related to unique crowds, this reputation may not be always relevant or not so relevant. Therefore, we wanted to check the validity of the argument above about Israeli related topics and non-Israeli related topics.

The question is: What is the measure of Israeli related articles in the Hebrew Wikipedia compare to non-Israeli related articles?

To answer this question, we have to determine criteria of quality and their weights.

The criteria we have chosen are listed below:

- Page size – page size in bytes.
- Total edits – the number of edits done to this article.
- Editors – the number of different editors.
- Page watchers – the number of different watchers.
- Pageviews – the number of views in last 60 days.
- Average time between edits - the average time passed between two consecutive edits (in days).
- Edits made by the top 10% of editors – number of edits made by 10% of the most influential editors regarding this page.
- Links to this page – number of internal links from Wikipedia lead to this page.
- Links from this page – number of internal links from Wikipedia this page leads to.
- External links – number of external links occur in the article.

These criteria were chosen due to multiple reasons:

1. There might be a positive correlation between the size of the page and the issues discussed in it.
2. Edits and editors are positively correlated with the amount of criticism this article has.
3. Views may point on the amount of popularity this article has, while this popularity might tell about this article's quality.
4. The links occur in the page show how other people relate to this article as a source of authority and what is the source dependent of this page.

The weights might be arguable, therefore we allowed the user to determine the weights by himself.

The reason we check both Israeli and non-Israeli related articles is that we are interested in the relation between the quality of Israeli and non-Israeli articles. For this reason it's not enough to check only the relation between Israeli articles and their corresponding articles in English.

2. First Experiment:

We ran our program over 883 articles that represent typical elements in Hebrew / Israel tradition, culture and history (HaShomer Organization, the Aliyah from Romania etc.) and investigate the results.

Over the 883, about 610 articles were also available in English (about 270 articles were not!) and the final measured ratio was 1.426117653, which means that the Hebrew articles are “better” by approximately 1.5 than the corresponding English articles.

3. The program as a generic tool:

As you can see below, there is an option to use the tool as a generic tool for examination of the relations between Hebrew and English articles in Wikipedia, though we designated it to check a specific question related to Israel.

Technically, in order to do this, we need to put list of articles' names in a file and activate the program on the second option (second radio-button).

To show this, we did a second experiment, as will be shown immediately.

4. Second Experiment:

Introduction:

It is known among the editors' community of the Hebrew Wikipedia, that there is a decision not to write about pornography topics as much as possible. As an example, In the English Wikipedia (and a few more languages) there is page contains a list of names of pornographic models, but in Hebrew there is not such page at all. We crawled over this page and extracted the names of articles in English which have an equivalent article in Hebrew. We got only 19 names in Hebrew from over 700 names in English (we note that the ratio is 35.58 in favor to the English Wikipedia), which is a very surprising fact! We executed the program on these 19 articles, using “plain reader weights”, and we got a ratio of: 1.4827836880050325.

These results (and the existence of much less articles in Hebrew in this topic) show that the decision of Hebrew Wikipedia is indeed imposed. It's not trivial that the editors indeed heed this order because people do it voluntarily.

In our opinion, these results are important and favorable because of the following reasons:

As can be seen in the list of names mentioned above, as time passed the number of females increased non-proportionally as opposed to the number of males. In addition, the fee of the females is much higher than the males' fee. The fundamental reason is that this industry based on objectification of the female gender and chauvinism (though there are people who interpret this state in positive way, due to the financial success and fame women get in contrast of the men, but we don't think that way). Due to this and to our equitable-feministic perspective we take kindly the fact that the editors obey this rule.