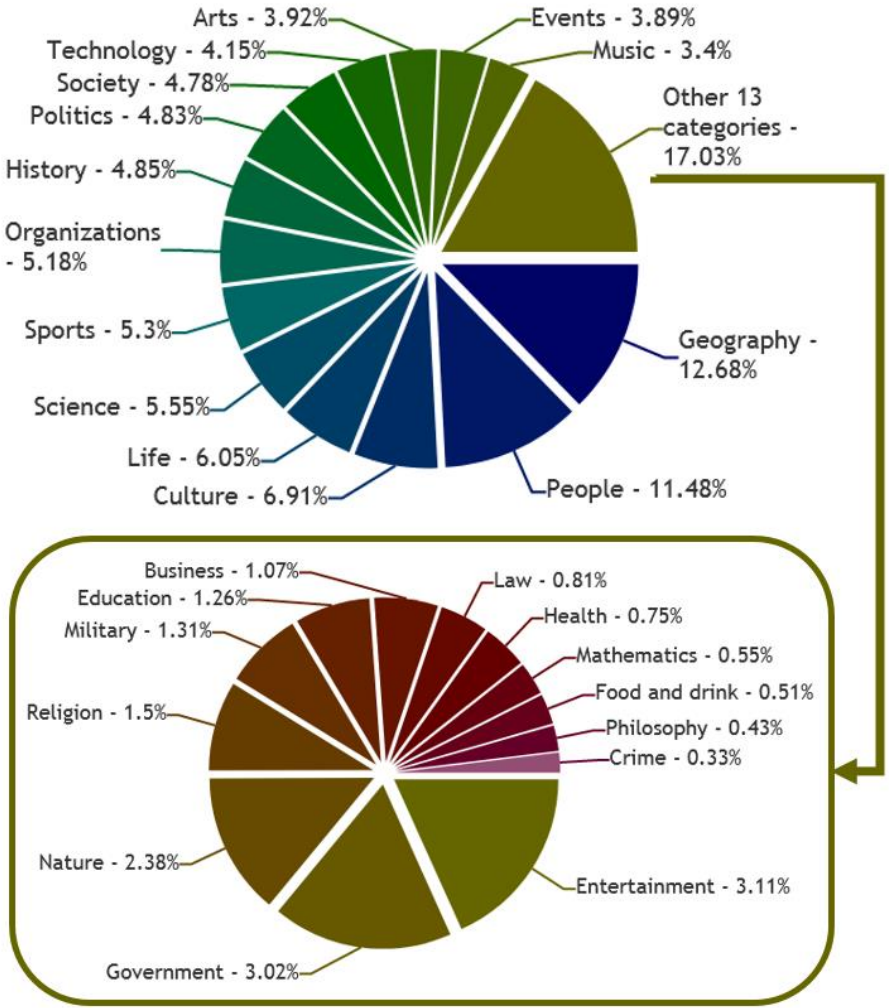
Multilingual Ranking of Wikipedia Articles

with Quality and Popularity Assessment in Different Topics

Multilingual Wikipedia

- Quality and popularity of information about the same topic depends on language
- Popularity can be measured from reader’ and authors’ point of view
- To assign articles to various topics it is possible to use different approaches
- There are over 10 million categories in Wikipedia



Categories

To classify the articles to one of the 27 main categories we took into account over 400 million links from articles to over 10 million categories and over 26 million links between categories in 55 Wikipedia language versions.

Figure presents resents how over 39 million Wikipedia articles distributed among 27 main categories.

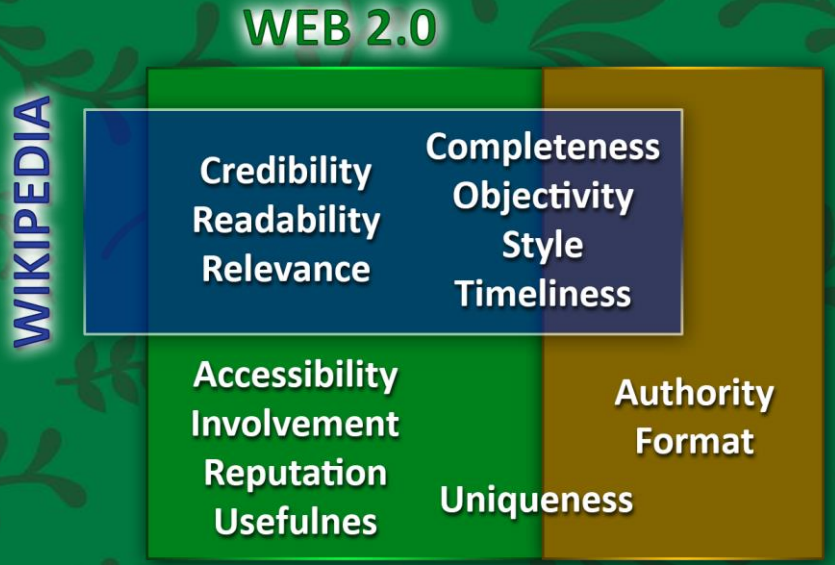
Source: https://doi.org/10.20944/preprints201905.0144.v2

WIKIMANIA STOCKHOLM

Quality

- Wikipedia quality dimensions overlap with some of traditional encyclopedia and Web 2.0 documents:
- completeness, credibility, objectivity, readability, relevance, style, timeliness.

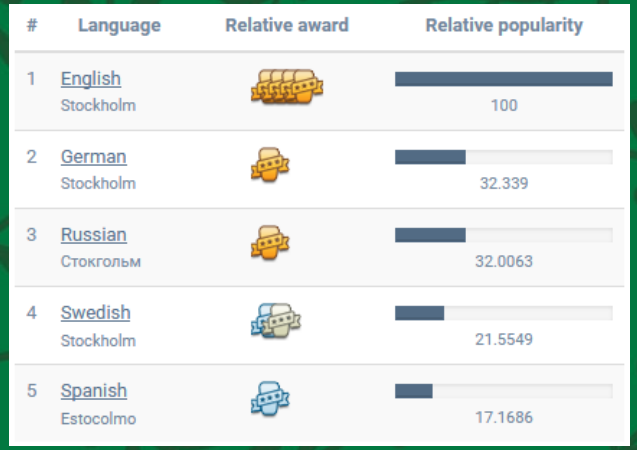
Source: https://doi.org/10.20944/preprints201905.0144.v2



Popularity

- Measures reflecting the demand for information contained in it by the readers and Wikipedia authors.
- Based on these measurements, we can compare not only different articles between each other, but also the language versions of the selected article

Source: https://wikirank.net/en/Stockholm



#	Language	Relative award	Relative popularity
1	English Stockholm		100
2	German Stockholm		32.339
3	Russian Стокгольм		32.0063
4	Swedish Stockholm		21.5549
5	Spanish Estocolmo		17.1686

#	Language	Relative award	Relative AI
1	Swedish Stockholm		100
2	English Stockholm		80
3	Russian Стокгольм		30
4	Esperanto Stokholmo		20
5	Portuguese Estocolmo		20

WIKIMANIA STOCKHOLM

Local & Global measures

$$PopGlobal(article) = \sum_{lang=1}^{n} PopLocal_{lang}(article),$$

where *PopLocal* means local popularity of the article, *lang* is the index of specific language version and *n* is number of the language versions of the selected article.

$$GlobalAI(article) = \left| \bigcup_{lang=1}^{n} Authors_{lang}(article) \right|,$$

where *Authors* means a set of authors' names, *lang* is the index of specific language version and *n* is the number of language versions of the article.

Source: https://doi.org/10.20944/preprints201905.0144.v2

WIKIMANIA STOCKHOLM

Assessment

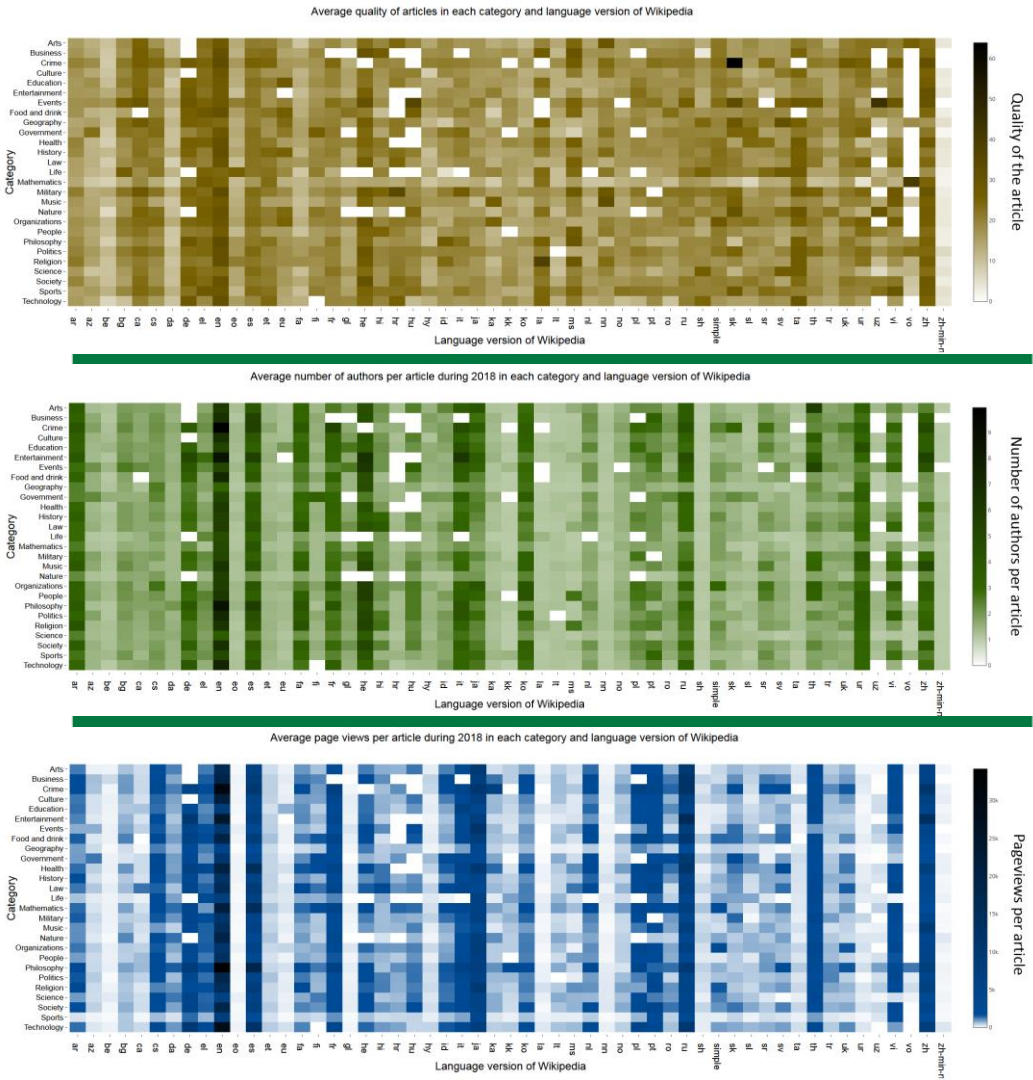
We extracted over 100 million values of features characterizing articles in all analyzed languages. Articles were grouped to 27 main categories.

These values were then used to calculate the measures:

- quality (**yellow heat map**),
- authors' interest (**green heat map**),
- popularity (**blue heat map**).

Source: https://doi.org/10.20944/preprints201905.0144.v2

WIKIMANIA STOCKHOLM



WikiRank

- Local rankings of Wikipedia articles in different languages.
- Global rankings, including specific topics.
- Quality and popularity measures of each of over 39 million Wikipedia articles.

WIKIMANIA STOCKHOLM

Related info

- Measures for Quality Assessment of Articles and Infoboxes in Multilingual Wikipedia (2019)
- Enrichment of multilingual Wikipedia based on quality analysis (Wikimedia CEE Meeting 2018)
- Relative Quality and Popularity Evaluation of Multilingual Wikipedia Articles (2017)

THANK YOU!