**Literature Review**

Keywords: Information Quality, Wikipedia, Machine Learning, Quality Metrics

**Objective**

Examine state of the art of existing metrics to measure Wikipedia article quality, and the usage of machine learning to predict it.

**Research Questions**

Q1: What are the most common metrics used to evaluate article quality in Wikipedia?

Q2: To what extent are those metrics currently used, and how effective are they?

Q3: Which machine learning approaches can be applied to predict article quality?

**Search Queries:**

* Google Scholar: Wikipedia (measuring OR measure) article quality
* ACM Digital Library: [Title: wikipedia assessing quality] AND [[Abstract: metric] OR [Abstract: measure]] AND [Full Text: wikipedia] AND [Full Text: assessing] AND [Full Text: quality] AND [Publication Date: (01/01/2000 TO 12/31/2021)]
* wikipedia assessing quality (metric OR measure)

All searches were restricted to a date range of 2000-2021.

**Inclusion Criteria**

I1: Papers discussing possible metrics to assess data quality in a collaborative encyclopedia.

I2: Papers discussing their findings and evaluating the metrics’ effectiveness.

I3: Papers discussing different machine learning approaches to predict article quality.

**Exclusion Criteria**

E1: Papers discussing vandalism and quality of user edits instead of the articles themselves.

E2: Papers discussing manual approaches to assess article quality

E3: Papers in a language other than English

**Process**

P1: Title is assessed for possible relevance, to be screened.

P2: Remove duplicates. (how many)

P3: Abstract too then, focused on inclusion and exclusion criteria

P4: Analyze full text with focus on research questions.

P5: Quality criteria?

Vandalismo 🡪 motivaçao

Que algoritmos usaram?

Posso usar outros artigos dps

Livros de machine learning: Ja deste

Pag. 8 -> tabela de números

Pag. 9 -> Medir qualidade de publicação

Estrategias de leitura? Definir, tipo procurar métricas, ou machine learning tipo geralmente lido em diagonal, e so dps ler mt a fundo.

Analise Wikipedia e métricas

Uso de estrategias machine learning

Table

Description automatically generated

**Google Scholar**

Advanced Search options are very limited (e.g, can’t combine title search with abstract/article search). As said in (https://www.mdpi.com/2073-8994/10/10/470), “The review was conducted by three reviewers and the search in all databases returned 6375 results. Due to a lack of advanced search options in Google Scholar and SpringerLink, they returned many non-related results. Therefore, we only included the first 200 most relevant results from these two databases”

Query: Wikipedia (measuring OR measure) article quality

Date range: 2000-2021

<https://scholar.google.com/scholar?q=Wikipedia+%28measuring+or+measure%29+AROUND%283%29+article+AROUND+%283%29+quality&hl=en&as_sdt=0%2C5&as_ylo=2000&as_yhi=2021>

Results: 18300, many irrelevant, so only screened the first 30 pages (300). Fetched about 50 papers that appeared relevant based on title and short description.

**ACM**

Query: [Title: wikipedia assessing quality] AND [[Abstract: metric] OR [Abstract: measure]] AND [Full Text: wikipedia] AND [Full Text: assessing] AND [Full Text: quality] AND [Publication Date: (01/01/2000 TO 12/31/2021)]

Date range: 2000-2021

https://dl.acm.org/action/doSearch?fillQuickSearch=false&target=advanced&expand=dl&AfterMonth=1&AfterYear=2000&BeforeMonth=12&BeforeYear=2021&AllField=Title%3A%28wikipedia+assessing+quality%29+AND+Abstract%3A%28metric+OR+measure%29+AND+Fulltext%3A%28wikipedia+AND+assessing+AND+quality%29

Results: 99

Fetched: 24

**Web of Science**

Query: wikipedia assessing quality (metric OR measure)

Date: 2000-2021

[**https://www.webofscience.com/wos/woscc/summary/9c126beb-2b28-4fd7-9f5b-a6862c992b25-17f692c7/relevance/1**](https://www.webofscience.com/wos/woscc/summary/9c126beb-2b28-4fd7-9f5b-a6862c992b25-17f692c7/relevance/1)

Results: 121

Fetched: 17

**Total: (Soma)**

**After Duplicates Removed: 76**

**Useful**

https://library.acg.edu/how-to-guides/google-scholar/advanced-searching

Extensive literature review example: <https://www.mdpi.com/2073-8994/10/10/470>

Important Tables from it:Table

Description automatically generated

Diagram

Description automatically generated