



dr Włodzimierz
Lewoniewski

Jego zainteresowania badawcze koncentrują się wokół jakości informacji w otwartych bazach wiedzy (takich jak Wikipedia, DBpedia, Wikidane), wykrywania fałszywych wiadomości, przetwarzania języka naturalnego, uczenia maszynowego.

wlodzimierz.lewoniewski
@ue.poznan.pl

JAK OCENIAĆ TREŚCI Z WIKIPEDII?

Wikipedia jest jedną z najpopularniejszych encyklopedii, swój sukces zawdzięcza autorom, czyli nam wszystkim. Czy zatem informacje w niej zawarte są wiarygodne?

Włodzimierz Lewoniewski

Katedra Informatyki Ekonomicznej
Uniwersytet Ekonomiczny w Poznaniu

Otwarty charakter internetu pozwala na poznanie się z różnorodnymi punktami widzenia i opiniami. Obecnie istnieje ponad miliard witryn internetowych, które oferują informacje na różny temat. Od wielu lat jednym z najbardziej cenionych serwisów internetowych jest Wikipedia. Ta ogólnodostępna encyklopedia zawiera obecnie ponad 60 mln artykułów w ponad 300 wersjach językowych. Umożliwia to dostęp do wiedzy dla ludzi z różnych kultur i regionów świata. Dodatkowo wiele popularnych serwisów internetowych (w tym wyszukiwarka Google i ChatGPT) korzystają z treści pochodzących z Wikipedii, by polepszać jakość swoich usług.

Wikipedia opiera się na modelu otwartej edycji, co oznacza, że każdy może tworzyć i zmieniać treści artykułów w ramach tej platformy. W związku z tym

informacje mogą być aktualizowane niemal natychmiast, serwis reaguje na bieżące wydarzenia i odkrycia. Wspólne edytowanie artykułów przez społeczność użytkowników (w tym anonimowych) pozwala na korektę błędów i ciągle doskonalenie treści.

Należy jednak zwrócić uwagę również na wady otwartego współredagowania Wikipedii. Na przykład artykuły tej encyklopedii są podatne na celowe wprowadzanie fałszywych informacji lub uszkodzenie treści. Ze względu na brak obowiązku sprawdzenia każdej edycji dokonanej przez dowolnego użytkownika te szkodliwe zmiany mogą być widoczne natychmiast dla innych czytelników Wikipedii. To prowadzi do błędów i nieścisłości w niektórych tekstach. Ponadto artykuły tej encyklopedii mogą być czasami stronnicze, zwłaszcza jeśli są edytowane przez osoby mające swoje interesy w określonym temacie. Należy również pamiętać, że możliwość edycji przez każdego może prowadzić do konfliktów między edytorami o treści artykułów lub interpretację zasad.

Mimo wyżej wymienionych wad ogólna filozofia umożliwiająca każdemu edycję Wikipedii była i jest kluczem do jej globalnego sukcesu. Artykuły na tej platformie są współtworzone przez wolontariuszy z różnych zakątków świata, co nadaje jej dynamikę i pozwala na bieżące aktualizacje. Każdego dnia w Wikipedii wprowadza się ponad pół miliona edycji, co sprawia, że ręczne monitorowanie wszystkich zmian jest trudne.

Artykuł na medal

Wikipedia w każdym języku jest tworzona przez unikatową społeczność użytkowników, która kształtuje i interpretuje standardy jakości dla swojej wersji językowej. Dlatego każda z nich ma nieco odmienne kryteria dotyczące jakości treści, wynikające z dyskusji i doświadczeń społeczności. Co do zasady każda wersja językowa ma specjalne wyróżnienia dla artykułów, które osiągnęły najwyższe standardy jakości.

Nazwy klas jakości w sześciu wersjach językowych Wikipedii: białoruskiej (be), niemieckiej (de), angielskiej (en), polskiej (pl), rosyjskiej (ru), ukraińskiej (uk)

| Ocena / Język | be | de | en | pl | ru | uk |
|--------------------|----|----|----|----|----|----|
| Artykuł na medal | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Klasa A | | | ✓ | | | |
| Dobry artykuł | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Solidny artykuł | | | | | ✓ | |
| Klasa B | | | ✓ | | | |
| Czwórka | | | | ✓ | | |
| Pełny artykuł | | | | | ✓ | ✓ |
| Klasa C | | | ✓ | | | |
| Rozwinięty artykuł | | | | | ✓ | ✓ |
| Klasa startowa | | | ✓ | ✓ | | |
| Artykuł w rozwoju | | | | | ✓ | ✓ |
| Załączek | ✓ | | ✓ | ✓ | ✓ | ✓ |



CLICKS/MEXICO/SHUTTERSTOCK.COM

W najbardziej rozwiniętej spośród wszystkich wersji językowych Wikipedii – anglojęzycznej – mianem *featured article* (FA) określa się wzorcowe artykuły, czyli te, które spełniają wszystkie kryteria jakości w danej wersji językowej. Z kolei *good article* (GA) to tytuł dla tekstów, które są bliskie standardom wzorcowych, lecz jeszcze ich w 100 proc. nie spełniają. W wersji polskojęzycznej Wikipedii takie treści są określane jako „artykuły na medal” i „dobre artykuły” (analogicznie do anglojęzycznych FA i GA).

Żeby artykuł mógł otrzymać jedno z tych prestiżowych wyróżnień, jest poddawany gruntownej analizie przez społeczność. Użytkownicy w ramach otwartych dyskusji decydują, czy dany tekst spełnia ustalone kryteria. Każdy może przedstawić argumenty za i przeciw jego uhonorowaniu. Warto dodać, że zasady przyznawania wyróżnień mogą ewoluować i bywają dostosowywane do potrzeb konkretnej wersji językowej. Takie zmiany mogą prowadzić do sytuacji, w której niektóre artykuły tracą wcześniej przyznaną im pozycję.

W niektórych wersjach językowych Wikipedii istnieje system oceny jakości artykułów, który wskazuje, jak blisko do osiągnięcia statusu wzorcowego jest dany artykuł. W anglojęzycznej Wikipedii artykuły są klasyfikowane w siedmiu kategoriach jakości, począwszy od najwyższej: FA, GA, A-class, B-class, C-class, Start, Stub. Interesujące jest to, że oceny niższe niż FA i GA są nadawane indywidualnie przez użytkowników bez konieczności prowadzenia dyskusji czy osiągania zgody społeczności. W polskiej Wikipedii poza najwyższymi wyróżnieniami artykuły mogą być klasyfikowane jako: „czwórka”, „start” lub „załążek”. Co ważne, w ramach tej samej wersji językowej w różnych dzia-

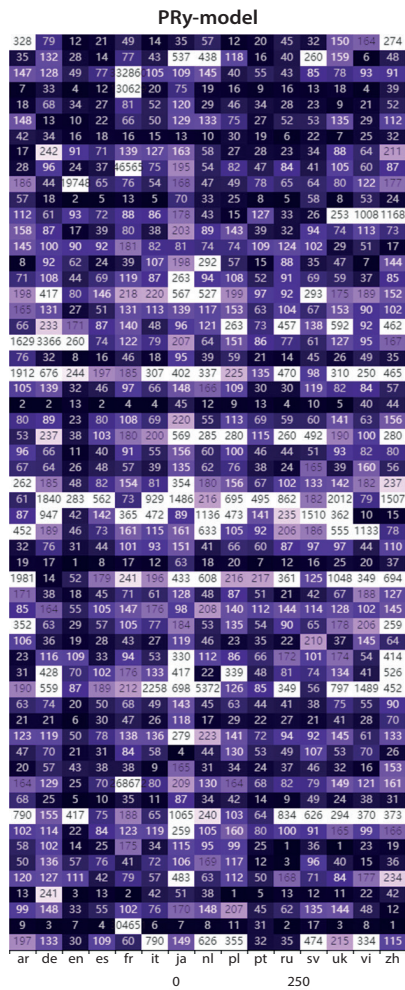
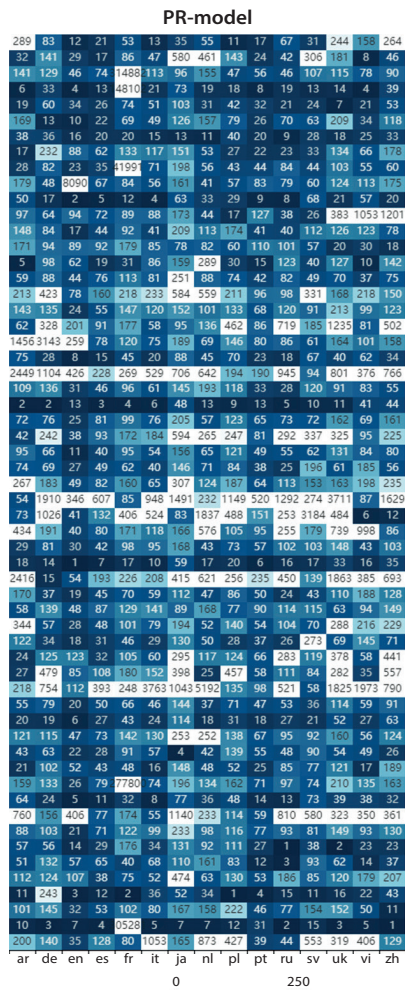
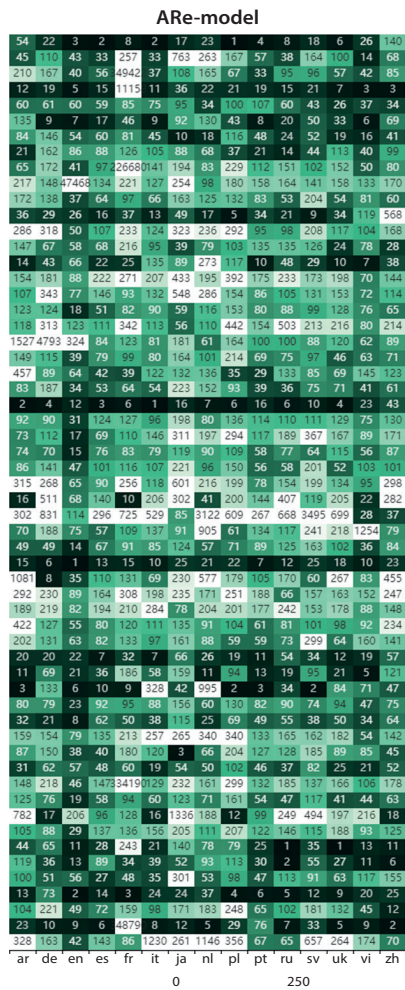
łach tematycznych mogą obowiązywać różne nazewnictwa dla podobnych poziomów jakości. Na przykład w pewnych obszarach polskiej Wikipedii klasa „czwórka” może być nazywana „poprawny”, a odpowiednikiem klasy startowej w innych działach jest klasa dostateczna.

Chociaż Wikipedia ma swoje standardy oceny jakości, to kryteria te mogą się różnić w zależności od wersji językowej i mogą się zmieniać. Ocenianie jest często subiektywne, co wymaga współpracy i uzgodnień między edytorami.

Wprowadzenie automatyzacji w proces oceny jakości treści na Wikipedii może przyczynić się do poprawy wiarygodności i skuteczności edycji. Algorytmy działają obiektywnie, bez emocjonalnych i subiektywnych interferencji, co może prowadzić do bardziej spójnej oceny. Pozwala to także na zbieranie danych na temat jakości i szybsze wykrywanie problemów.

Narzędzia informatyczne mogą szybko identyfikować wandalizmy czy dezinformacje, dostarczając edytorom bieżących informacji i sugerując poprawki. Dodatkowo, w przypadku ataków dezinformacyjnych, te narzędzia działają jako mechanizm obronny.

Katedra Informatyki Ekonomicznej na Uniwersytecie Ekonomicznym w Poznaniu prowadzi badania oceny jakości artykułów Wikipedii. Niektóre modele opracowane w ich ramach oraz opisane w publikacjach naukowych zostały zaimplementowane w niektórych ogólnodostępnych serwisach internetowych. Na przykład projekt WikiRank.net pozwala na sprawdzenie oceny jakości oraz popularności artykułów w różnych wersjach językowych. Inny przykład – projekt do analizy ważności źródeł informacji BestRef.net.



LEWONIEWSKI (2022)

zapobiegając ich rozprzestrzenianiu się. Ponadto nowi redaktorzy Wikipedii mogą nie być pewni, jakie źródła są najbardziej wiarygodne w danej dziedzinie. Automatyczna ocena źródeł może dostarczyć im wskazówek i zaleceń, pomagając w wyborze odpowiednich materiałów źródłowych.

Badanie wszystkich artykułów Wikipedii w różnych wersjach językowych pokazało, że istnieje ponad milion różnych wytrny internetowych znajdujących się w ponad 300 mln przypisów (referencjach). W najbardziej rozwiniętej wersji językowej (angielskiej) można zidentyfikować blisko 77 mln przypisów, w wersji polskojęzycznej – około 8 mln. Wykorzystując różne modele oceny źródeł internetowych, można zidentyfikować te najistotniejsze z punktu widzenia poszczególnej wersji językowej Wikipedii.

Semantyczne bazy danych

Rozwój technologii semantycznych znacząco poprawia efektywność takich procesów jak wyszukiwanie informacji, analiza sentymentu czy automatyczne podsumowanie treści. Dobrymi przykładami otwar-

tych i powiązanych z Wikipedią semantycznych baz wiedzy są Wikidane i DBpedia.

Podczas gdy DBpedia koncentruje się na wyodrębnianiu danych z Wikipedii w formie bardziej przystępnej dla maszyn, Wikidane służą jako centralna baza danych wspierająca wszystkie projekty Wikimedia w różnych językach. Wspólnie te inicjatywy przyczyniają się do zwiększenia dostępu do wiedzy w sposób bardziej strukturyzowany. Poprawa jakości w Wikipedii może również przyczynić się do polepszenia tych semantycznych baz wiedzy.

Wikipedia, Wikidane i DBpedia są otwartymi zasobami, które umożliwiają wykorzystanie swojej treści do różnorodnych celów. Lepsza jakość tych zasobów może się przyczynić do polepszenia innych serwisów, które wykorzystują otwarte dane. Poniżej znajduje się lista przykładów serwisów i zastosowań, które mogą korzystać z Wikipedii, Wikidany i DBpedia: wyszukiwarki internetowe, przetwarzanie języka naturalnego, aplikacje edukacyjne, systemy rekomendacji, asystenci wirtualni i chatboty, aplikacje kulturowe i turystyczne, analiza sieci powiązanych oraz wiele innych. ■

Chcesz wiedzieć więcej?

Lewoniewski W., Węcel K., Abramowicz W., *Companies in Multilingual Wikipedia: Articles Quality and Important Sources of Information*, w: *Lecture Notes in Business Information Processing*, vol. 471, 2023.

Lewoniewski W., *Identification of Important Web Sources of Information on Wikipedia across various Topics and Languages*, „*Procedia Computer Science*” 2022, vol. 207.

Lewoniewski W., Węcel K., Abramowicz W., *Reliability in Time: Evaluating the Web Sources of Information on COVID-19 in Wikipedia across Various Language Editions from the Beginning of the Pandemic*, 2022.

https://www.wbc.poznan.pl/Content/461699/Lewoniewski_Wlodzimierz-rozprawa_doktorska.pdf