

Abstract

Models and support methods of the Wikipedia knowledge Community

Wikipedia is a free online encyclopedia consisting of over 47 million articles in 288 different languages. It was founded on 15th January 2001 and is currently one of the most popular websites in Alexa rating. Wikipedia is one of the greatest examples of how the online community can collaboratively produce online content. While Wikipedia is considered a mature Web2.0 site and used as a model for other Web2.0 sites, it is still facing numerous constantly emerging and unpredictable problems, such as: decline in growth of the English-language Wikipedia; criticism of Wikipedia content quality, credibility or controversy, or a decline of successful RfA procedure.

New technology is needed to support the continuous quality improvement efforts of the Wikipedia knowledge community. Moreover, such new methods could be applied to other projects that use the Wiki technology, including commercial enterprises, in the future.

One of the main goals of this dissertation is to support decisions and decision processes of the Wikipedia knowledge community that aim to improve article quality or decision making among editors. Such decision support technology requires good models of the Wikipedia knowledge community. The proposed new methods and, in particular, the automatic article quality evaluation method have a satisfying performance. Therefore, they can be used for developing tools that can help in solving practical problems of the Wikipedia knowledge community, such as Wikipedia article quality evaluation or admin candidate selection. The proposed methods are based on a new model of the Wikipedia knowledge community, expressed using a Multidimensional Behavioral Social Network (MBSN).

Furthermore, many of the most important attributes used for building those models can be extracted and calculated for other online collaborative projects not related to the Wikipedia that use the Wiki technology. This shows the opportunity for generalization of the results and methods proposed in this thesis.

Finally, studies on RfA and new methods for predicting results of the RfA procedure show that it should be possible to improve the likelihood of electing new, good admins by changing the criteria of nomination for the RfA procedure. Research suggests that there could be two criteria: one based on the number of edits, the other based on the diversity of edits (measured by the number of topical categories in which a candidate edited articles).

Very important contribution of the thesis is model that was verified based on a comparison of behavioral social networks with declarative social networks created in a large online experiment. This is the first attempt to verify social network analysis results with actual answers

given during the surveys made by Wikipedia contributors. In addition, the results of the survey were used to improve the MBSN model. The results of these studies were published in the dissertation and in the prestigious journal Information Sciences.

Streszczenie

Models and support methods of the Wikipedia knowledge Community

Wikipedia jest wielojęzyczną darmową encyklopedią internetową działającą w oparciu o zasadę otwartej treści. Zawiera ponad 47 milionów wpisów w 288 różnych językach. Została założona w styczniu 2001, a obecnie jest jedną z najpopularniejszych stron internetowych według rankingu Alexa. Wikipedia jest doskonałym przykładem sposobu, w jaki rozproszona społeczność użytkowników Internetu może wspólnie pracować nad treścią. Pomimo tego, że Wikipedia jest uznawana za dojrzały projekt Web 2.0 i podawana za wzór dla innych projektów, nadal zmaga się z wieloma nowo powstającymi problemami takimi jak: spowolnieniem wzrostu angielskiej wersji językowej; problemem jakości artykułów i ich kontrowersyjności; mniejszą skutecznością wyboru nowych administratorów za pomocą procedury RfA. Z tych powodów potrzebne są nowe metody, narzędzia lub procedury wspomagające problematyczne procesy. Dodatkowo, podobne rozwiązania będą mogły znaleźć zastosowanie w innych społecznościach Internetowej współpracy.

W niniejszej rozprawie głównym celem jest opracowanie metody wspierającej proces oceny jakości artykułów. Do tego celu wymagane jest stworzenie dobrych modeli reprezentujących wiedzę i relacje pomiędzy twórcami w społeczności Wikipedii. W pracy proponowana jest metoda automatycznej oceny jakości artykułów posiadająca satysfakcjonującą dokładność. Metoda ta może być następnie użyta do budowy narzędzi i wsparcia procedur oceny jakości treści lub wyborów administratorów systemu. Zaproponowana metoda opiera się na użyciu wielowymiarowej społecznej sieci behawioralnej (Multidimensional Behavioral Social Network – MBSN). Ponadto, wiele najistotniejszych atrybutów użytych do budowania sieci i modeli może zostać wydobyta i obliczona dla innych niż Wikipedia społeczności online. Udowadnia to potencjał do uogólnienia wyników badań niniejszej rozprawy. Ostatnia część pracy dotyczy poprawy jakości procedury wyboru administratorów i stworzenia mechanizmu automatycznego sugerowania najlepszych do tej roli użytkowników Wikipedii. Badania wskazują, że mogą być użyte do tego celu dwa główne kryteria: pierwsze na podstawie ogólnej liczby edycji i aktywności użytkownika, oraz drugie na podstawie różnorodności tematycznej artykułów edytowanych przez użytkownika.

Ważnym osiągnięciem rozprawy jest weryfikacja modelu społeczności autorów Wikipedii, jakim jest sieć MBSN, za pomocą sondażu. Wyniki sondażu dotyczącego relacji pomiędzy autorami Wikipedii zostały porównane z behawioralnymi sieciami. Dodatkowo, wyniki sondażu zostały

użyte do udoskonalenia modelu MBSN. Wyniki tych badań zostały opublikowane w rozprawie i w prestiżowym czasopiśmie Information Sciences.