Excavating the mother lode of human-generated text: A systematic review of research that uses the Wikipedia corpus - DTU Orbit (18/01/2019)

Excavating the mother lode of human-generated text: A systematic review of research that uses the Wikipedia corpus

Although primarily an encyclopedia, Wikipedia's expansive content provides a knowledge base that has been continuously exploited by researchers in a wide variety of domains. This article systematically reviews the scholarly studies that have used Wikipedia as a data source, and investigates the means by which Wikipedia has been employed in three main computer science research areas: information retrieval, natural language processing, and ontology building. We report and discuss the research trends of the identified and examined studies. We further identify and classify a list of tools that can be used to extract data from Wikipedia, and compile a list of currently available data sets extracted from Wikipedia.

General information
State: Published
Organisations: Department of Applied Mathematics and Computer Science, Cognitive Systems, Elon University, Concordia University, University of Oulu
Contributors: Mehdi, M., Okoli, C., Mesgari, M., Nielsen, F. Å., Lanamäki, A.
Number of pages: 25
Pages: 505–529
Publication date: 2017
Peer-reviewed: Yes

Publication information
Journal: Information Processing & Management
Volume: 53
Issue number: 2
ISSN (Print): 0306-4573
Ratings:
BFI (2019): BFI-level 2
Web of Science (2019): Indexed yes
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): CiteScore 4.23 SJR 0.92 SNIP 2.66
Web of Science (2017): Impact factor 3.444
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 2.83 SJR 0.705 SNIP 2.007
Web of Science (2016): Impact factor 2.391
BFI (2015): BFI-level 2
Scopus rating (2015): CiteScore 2.37 SJR 0.73 SNIP 1.978
Web of Science (2015): Impact factor 1.397
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 2.07 SJR 0.762 SNIP 1.955
Web of Science (2014): Impact factor 1.265
BFI (2013): BFI-level 2
Scopus rating (2013): CiteScore 2.05 SJR 0.675 SNIP 1.614
Web of Science (2013): Impact factor 1.069
ISI indexed (2013): ISI indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): CiteScore 1.99 SJR 0.606 SNIP 1.684
Web of Science (2012): Impact factor 0.817
ISI indexed (2012): ISI indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): CiteScore 2.58 SJR 0.849 SNIP 2.588
Web of Science (2011): Impact factor 1.119
ISI indexed (2011): ISI indexed yes
BFI (2010): BFI-level 2
Scopus rating (2010): SJR 0.877 SNIP 2.198