Spatio-temporal Analysis of Reverted Wikipedia Edits

Johannes Kiesel, Martin Potthast, Matthias Hagen, Benno Stein
<first name>.<last name>@uni-weimar.de
Bauhaus-Universität Weimar
www.webis.de

ICWSM-17, May 18th 2017
Plato
From Wikipedia, the free encyclopedia

Plato (ancient Greek: Πλάτων, Plátōn, "wide, broad-shouldered") (c. 428/427 BC[8] - c. 348/347 BC) was an ancient Greek philosopher, the second of the great trio of ancient Greeks: Socrates, Plato, and Aristotle— who between them laid the philosophical foundations of Western culture.[1] Plato was also a mathematician, writer of philosophical dialogues, and founder of the Academy in Athens, the first institution of higher learning in the western world. [Verification needed] Plato is widely believed to have been a student of Socrates and to have been deeply influenced by his teacher's unjust death.

Plato's brilliance as a writer and thinker can be witnessed by reading his Socratic dialogues. Some of the dialogues, letters, and other works that are ascribed to him are considered spurious.[2] However, it is very probable that "everything Plato wrote for publication" has been passed along to us intact.[3]
What is Vandalism in Wikipedia?

Plato

From Wikipedia, the free encyclopedia

Plato (ancient Greek: Πλάτων, Plátōn, "wide, broad-shouldered") (c. 428/427 BC[1] – c. 348/347 BC) was an ancient Greek philosopher and one of the great trio of ancient Greeks – Socrates and Aristotle – who between them laid the philosophical foundations of Western culture.[1] Plato was also a mathematician, a poet, a dialogue writer, and a political theorist; his dialogues, the first institution of higher learning, are the most influential philosophical dialogues, and founded the Academy in Athens, the first institution of higher learning in the Western tradition.[2] [Verification needed] Plato is widely believed to have been a disciple of Socrates and to have been involved in his trial and death.[3]

Plato was a very influential philosopher and thinker. His ideas and theories have been studied and debated for centuries. His works have had a profound impact on Western philosophy and continue to influence modern thought. Plato's dialogues, in particular, have been studied extensively, and his ideas have been applied in various fields, including science, politics, and ethics. His influence can be seen in the works of many other philosophers, and his ideas continue to be relevant today.
Is Vandalism a Problem for Wikipedia?

- 40 million (9.5%) are vandalism
  → a vandalism case every 10s
Is Vandalism a Problem for Wikipedia?

- 40 million (9.5%) are vandalism
  → a vandalism case every 10s

Countermeasure: Bots that detect and revert vandalism

Problem: False positives of the bots discourage editors
Towards Understanding Vandalism in Wikipedia

How to avoid vandalism in the first place?

→ Understand **why** people vandalize Wikipedia.
→ Analyze **when** people vandalize.
→ Analyze **where** these people are.
Towards Understanding Vandalism in Wikipedia

How to avoid vandalism in the first place?

→ Understand why people vandalize Wikipedia.

→ Analyze when people vandalize.

→ Analyze where these people are.

We analyzed all 1.2 billion edits to the 7 most-edited Wikipedias

- Large-scale mining of vandalism using reverted edits
- Historical geolocation of anonymous editors by cross-checking several geolocation sources
- Spatio-temporal analysis revealing when anonymous editors vandalize from where
Mining Vandalism Using Reverted Edits
Mining Vandalism Using Reverted Edits

Article over time

Editor

Article revision

Edit
Mining Vandalism Using Reverted Edits

Article over time

Editor

Article revision

Edit
Mining Vandalism Using Reverted Edits

Article over time

Editor

Article revision

Edit

Revert

Reverted edits
Mining Vandalism Using Reverted Edits

- Not all reverted edits are vandalism
- Relying on non-obligatory revert comments underestimates vandalism

Identified 7 patterns of non-vandalism or ambiguous reverts

- Revert to blank page
- Empty revert
- Revert correction (enlargement)
- Revert reverting more than one editor
- Interleaved reverts (edit war)
- Reverted revert
- Self-revert

Filter 67% of reverted edits

Vandalism detection with precision 82.8%, recall 84.7%
Analyzing Vandalism in Wikipedia (by time)
Analyzing Vandalism in Wikipedia (by time)

Vandalism ratio = \frac{\text{vandalism edits}}{\text{edits}}
Analyzing Vandalism in Wikipedia (by time)

English Wikipedia from United States

- Monday - Thursday
- Friday
- Saturday
- Sunday

Estimates from less than 1000 vandalism edits are shown as dotted lines.

@KieselJohannes

@Bauhaus-Universität Weimar
Analyzing Vandalism in Wikipedia (by time)

Estimates from less than 1000 vandalism edits are shown as dotted lines.
Analyzing Vandalism in Wikipedia (by time)

Japanese Wikipedia from Japan

Estimates from less than 1000 vandalism edits are shown as dotted lines.

Monday - Thursday
Friday
Saturday
Sunday
Analyzing Vandalism in Wikipedia (by time)

French Wikipedia from France

Monday - Thursday
Friday
Saturday
Sunday

Estimates from less than 1000 vandalism edits are shown as dotted lines.

Wednesday
Analyzing Vandalism in Wikipedia (by country)

Country estimates from less than 1000 vandalism edits are not colored.
Analyzing Vandalism in Wikipedia (by country with English as an official language)

Country estimates from less than 1000 vandalism edits are not colored
Analyzing Vandalism in Wikipedia (by time)

English Wikipedia from Germany

German Wikipedia from Germany

Estimates from less than 1000 vandalism edits are shown as dotted lines

Monday - Thursday        Friday        Saturday        Sunday

@KieselJohannes
Analyzing Vandalism in Wikipedia (by time)

Estimates from less than 1000 vandalism edits are shown as dotted lines
Spatio-temporal Analysis of Reverted Wikipedia Edits

Future Work

- Identify different types of vandalism
- Identify changes in vandalism behavior over the years

Resources

- Interactive tool for exploring the vandalism ratio graphs
  [webis16.medien.uni-weimar.de/wikipedia-vandalism](webis16.medien.uni-weimar.de/wikipedia-vandalism)
- Supplementary material (~50 pages of tables and graphs)
  [github.com/webis-de/ICWSM-17/raw/master/supplementary-material.pdf](github.com/webis-de/ICWSM-17/raw/master/supplementary-material.pdf)
- Code for historical geolocation
  [github.com/webis-de/aitools4-aq-geolocation](github.com/webis-de/aitools4-aq-geolocation)
- Code for reproducing experiments
  [github.com/webis-de/ICWSM-17](github.com/webis-de/ICWSM-17)