Towards Vandalism Detection in Knowledge Bases: Corpus Construction and Analysis

Stefan Heindorf       Martin Potthast       Benno Stein       Gregor Engels

Motivation

Context

Problems

Solution Idea

Related work

Concentrates on unstructured knowledge bases

ML Machine learning to detect vandalism

Vandalism corpus

Contributions

• Wikidata Vandalism Corpus WDVC-2015
• Corpus analysis

Related work

Concentrates on unstructured knowledge bases

Corpus Construction

Automatic Revision Labeling

• Wikidata revision history
• Only non-bot revisions considered
• Goal: Automatic labeling as vandalism/non-vandalism

Option 1: Rollback (by administrators)
103,205 rollbacked revisions

Option 2: Undo/Restore (by all users)
64,820 undone/reverted revisions

Manual Validation

1,000 rollbacked revision
Manually reviewed 1,000 undone/reverted revisions
1,000 inconspicuous revisions

Option 1: Rollback (by administrators)
86 ± 3 %* revisions indeed vandalism

Option 2: Undo/Restore (by all users)
62 ± 3 %* revisions indeed vandalism

* 95 % confidence level

For our corpus: Rollbacked revisions are considered vandalism

Overview of Corpus

• 24 million revisions
• 103 thousand vandalism revisions

Corpus Revisions over Time

What is vandalized?

40% of vandalism in structured data

Who vandalizes?

86 % of vandalism by unregistered users

Top vandalized items

47 Cristiano Ronaldo
43 Lionel Messi
43 One Direction
41 Portal:Featured content
34 Justin Bieber
33 Barack Obama
29 English Wikipedia
29 Selena Gomez

Top items by category

Culture 12%
People 12%
Society 12%
Nature 12%
Meta items 12%
Technology 12%
Places 12%
Other 12%

Download our corpus!

SIGIR '15 - Santiago de Chile
heindorf@uni-paderborn.de