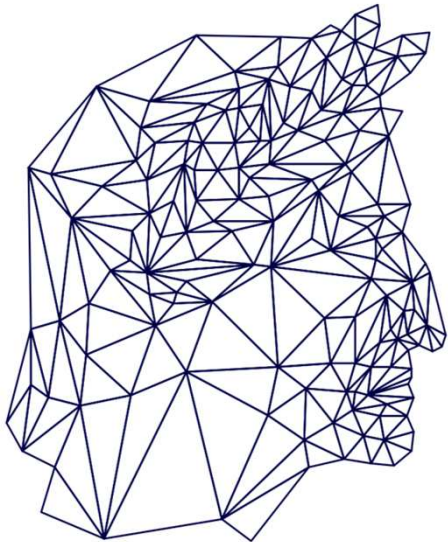


Tool AI vandalismi



itWikiCon
mo
2018

Cos'è un vandalismo? (1/3)

- Si definisce vandalismo l'aggiunta, la cancellazione o la modifica di contenuti e dati fatta con un evidente interesse o una malafede e con il conseguente risultato di compromettere l'integrità di Wikipedia.
- Viceversa, qualsiasi contributo fatto in buona fede finalizzato al miglioramento dell'enciclopedia, anche se fatto in modo errato o maldestro, non è da considerarsi vandalismo. Gli interventi nei quali non è ravvisabile in modo immediato e inequivocabile la malafede non sono considerati vandalismo.

(da it.Wikipedia:Vandalismo)

Cos'è un vandalismo? (2/3)

- *Spamming, come l'aggiunta di collegamenti esterni a fini promozionali o a siti non consentiti dalle regole di Wikipedia*
- *Aggiunta di parole senza senso o cancellazione non giustificata di parti del testo*
- *Creazione di voci prive di senso o con contenuto scherzoso/goliardico*
- *Inserimento, cancellazione o modifica di contenuti, template, dati, compiuti con un evidente interesse, una malafede e contrariamente al consenso, specie se reiteratamente*
- *Inserimento di informazioni volutamente errate*
- *Inserimento a fini di disturbo, anche in pagine di discussione, di testi in lingua diversa dall'italiano*

Cos'è un vandalismo? (3/3)

- *Inserimento di insulti o ingiurie, cancellazione, danneggiamento di pagine utente altrui*
- *Danneggiamento delle pagine di discussione delle voci di Wikipedia o degli utenti iscritti*
- *Inserimento di immagini provocatorie o volutamente fuori contesto*
- *Inserimento intenzionale di template non appropriati o eliminazione reiterata di template appropriati*
- *Spostamento di pagine in modo intenzionalmente inappropriato*

Cos'è l'I.A.?

- In generale: *«L'intelligenza artificiale (o IA, dalle iniziali delle due parole) è una disciplina appartenente all'informatica che studia i fondamenti teorici, le metodologie e le tecniche che consentono la progettazione di sistemi hardware e sistemi di programmi software capaci di fornire all'elaboratore elettronico prestazioni che, a un osservatore comune, sembrerebbero essere di pertinenza esclusiva dell'intelligenza umana.»* (Marco Somalvico)
- Nel particolare di questo strumento, è un insieme di programmi che, istruito attraverso esempi di giudizio (espressi da umani) sulla qualità degli edit, ambisce a sviluppare una logica decisionale che riesca a giudicare un futuro edit con lo stesso grado di attendibilità di un ipotetico utente esperto sia nell'argomento della voce sia nella lotta al vandalismo.

Come si fa oggi la lotta ai vandalismi in Wikipedia?

- *Uso di CAPTCHA per certe tipologie di edit per evitare l'uso di sistemi di inserimento automatico*
- *Blocco alla creazione di voci con titoli in «black list»*
- *Inserimento di tag nell'oggetto al momento del salvataggio*
- *Controllo della coda delle «ultime modifiche» con VandalFighter*
- *«Patrolling» sulle modifiche degli osservati speciali*

Perché serve l'aiuto dell'I.A.? (1/2)

- Il numero di voci di ogni wiki cresce nel tempo
- Il numero di utenti «attivi» tende a stabilizzarsi nel tempo
- La crescita di «notorietà» di wikipedia attira più vandali:
 - quelli spinti dal desiderio di dire «son riuscito a piazzare una bufala e ci han messo settimane prima di accorgesene»
 - quelli spinti dalla volontà di inserire contenuti non NPOV per finalità ideologiche o promozionali

Perché serve l'aiuto dell'I.A.? (2/2)

- In futuro, se si volesse mantenere la stessa capacità di contrasto ai vandalismi di oggi, ogni patroller dovrebbe:
 - occuparsi di un maggior numero di voci
 - sanare un maggior numero di vandalismi
 - saper identificare con la stessa rapidità le sofisticate alterazioni (ora più probabili che in passato) e le grezze goliardate
 - iniziare con il bagaglio di esperienza di chi lascia
- In altre parole: UN FALLIMENTO CERTO

Come funziona il progetto WMF per I.A?

- Il progetto prevede l'addestramento del sistema di I.A. su vari classi di modelli.
- Ciascun modello è ciò che guida il sistema di A.I. a esprimere un giudizio.
- Le classi ad oggi in addestramento sono:
 - «Edit quality»: servono a giudicare il singolo edit
 - «Article quality»: servono a giudicare la voce nel suo insieme
 - «Draft quality»: servono a giudicare le nuovi voci
 - «Draft topic»: servono a segnalare a utenti competenti per tema le nuovi voci

Come funziona il progetto WMF per I.A?

- L'addestramento avviene tramite la sottomissione di una serie di casi campione a utenti umani che esprimono il loro giudizio.
- Per ogni classe vi sono vari modelli che si focalizzano su elementi diversi, creando così un percorso di addestramento e evoluzione con l'obiettivo di aumentare progressivamente l'affidabilità del giudizio espresso dall'I.A.
- <https://labels.wmflabs.org/ui/itwiki/>

A che stadio è il progetto?

- Lo stadio di avanzamento del progetto è fortemente variabile rispetto all'edizione linguistica; ad esempio `{{en}}` vs `{{it}}`:

| | Edit Quality | Article Quality | Draft Quality | Draft Topic |
|---------------------|--------------|-----------------|---------------|-------------|
| <code>{{en}}</code> | Live | Live | Live | Training |
| <code>{{it}}</code> | Training* | - | - | - |

- Quanto il sistema di A.I. apprende con l'addestramento fatto in una specifica lingua può essere utilizzato solo per quella lingua.
- Senza l'impegno della comunità ad addestrare il sistema nelle varie classi, il progetto non può progredire per la lingua di quella comunità.

Come si può contribuire al progetto?

- Partecipando ai task di training
- Creare un support team italofono, con utenti attivi in quanti più possibili progetti tematici, per coadiuvare il team del progetto A.I. spiegandogli le metriche di merito usate in it.wiki, ad esempio per i criteri di «cancella subito», i criteri di assegnamento delle stellette alle voci di qualità.

Collegamento con il capo progetto WLM

Aaron Halfaker

Principal Research Scientist, Wikimedia Foundation

Think big. Measure what you can. Build better technologies.



[Halfaker, Aaron, Sept. 2013.jpg](#) (CC-BY-SA 3.0)

About me

Hi. I'm Aaron Halfaker. I'm a scientist. See [projects](#) and [publications](#) below. I've been a Wikipedian since 2008. I mostly build tools and run studies, but I make edits where I can. In 2011, I started working with the Wikimedia Foundation as a research scientist. This is my staff account. See my volunteer account, [User:EpochFail](#), for my non-staff work.

My work

My job is to build understanding about and support for the socio-technical fabric of the Wikimedia movement. I tend to focus on our computer mediated spaces (Wikipedia, Commons, Wikidata, Wikisource, etc.) and quality dynamics (patrolling, curation, newcomer training, etc.)

See [The Rise and Decline](#) for an example of my research and [Revision scoring](#) as a service for an example of an experimental technology I'm actively working on.

Contact me

- E-mail: ahalfaker@wikimedia.org
- Website: <http://halfaker.info/>
- Twitter: <http://twitter.com/halfaker/>
- On IRC, I go by "halfak".

iD

This user has the ORCID identifier:
0000-0001-8907-0307

[https://meta.wikimedia.org/wiki/User:Halfak_\(WME\)](https://meta.wikimedia.org/wiki/User:Halfak_(WME)) (CC-BY-SA 3.0)

Crediti

Slide 1: File:Logo itWikiCon 2018.png by Afnecors

Slide 13: File:Using_artificial_intelligence_to_keep_Wikipedia_open_(Wikimania_2018).pdf by EpochFail (pagina 2)