

Machine Learning as a Service na Wikipédia

Henrique de Andrade

RioInfo - 27 de setembro de 2017

Wikipédia

- 20 bilhões de pageviews por mês
- 500 mil usuários ativos
- 40 milhões de artigos
- 299 idiomas
- 1000 edições por minuto

Wikipédia

- Qualquer um pode editar
 - Todos podem revisar
mas dá um trabalho...

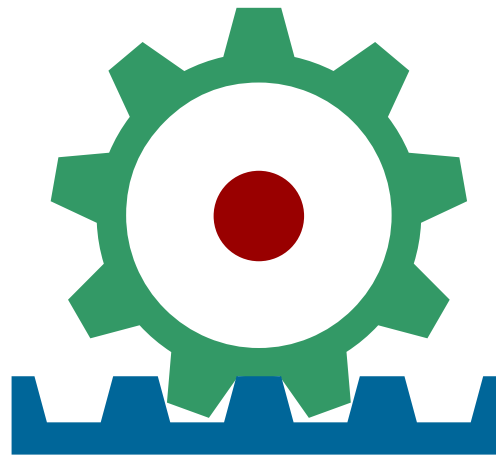
Ferramentas de apoio a edição

- Huggle
- Snuggle
- Cluebot NG



ORES

Objective Revision Evaluation Service



'framework revscoring

Modelos

- Qualidade da edição
 - reverted
 - damaging
 - goodfaith
- Qualidade do artigo
 - wp1.0
 - draftquality

32 projetos em 29 idiomas

Exemplo

edição 1 edição 2 edição 3

score: <https://ores.wikimedia.org/v3/scores/ptwiki/?models=damaging|goodfaith|reverted&revids=49801894|49958129>

ui: <https://ores.wikimedia.org/ui/>

Treinando o ORES

- Histórico de reversões
 - BWDS (tf-idf)
 - lista para o português
 - badwords
 - informals

Treinando o ORES

- goodfaith e damaging
 - Avaliação direta dos usuários

Wiki Labels



- campanhas correntes
 - campanha na Wikipédia em inglês

JADE

Judgement and Dialog Engine

Best practices for AI in the social spaces: Integrated refutations

- ferramenta em desenvolvimento
- hoje falsos positivos são identificados na mão

Usos atuais

- Ferramentas de monitoramento
 - Filtros em mudanças recentes
 - crosswatch
 - CopyPatrol
 - ra.un

reduzido 90% do tempo gasto revisando edições

Usos atuais

- Combate ao vandalismo
 - PatruBOT
- Sistema de Recomendação
 - SuggestBot
- Acompanhamento de qualidade
 - Reports Bot
 - Qualidade relativa de categorias
 - Programa de Educação

Saiba Mais



<http://iwsc.co>

IWSC



- Aaron Hafaker
 - Principal Research Scientist

Agradecimientos

- Adam Wight
- Amir Sarabadani
- Sumit
- とある白い猫
- Arthur Tilley
- He7d3r
- Yuvipanda



Obrigado!

henrique@wikibrasil.org



Wiki Educação
—— Brasil ——