

Usos científicos dos dados da Wikipédia

Jonas Xavier
jonas.agx@gmail.com

IFPE - Recife
Wikimedia Brasil

Fevereiro de 2012

Sumário

Wikipédia

Introdução

- Estrutura

- Conteúdo

- Aspectos sociais

Tendências sociais fora da web

Perguntas

Referências

Introdução

Os trabalhos acadêmicos sobre a Wikipédia podem ser divididos em dois grupos principais:

- ▶ Trabalhos sobre seu **conteúdo**, os artigos.
- ▶ Aspectos sociais da comunidade dos editores ou leitores, os usuários, como número de edições, mensagens.

Introdução

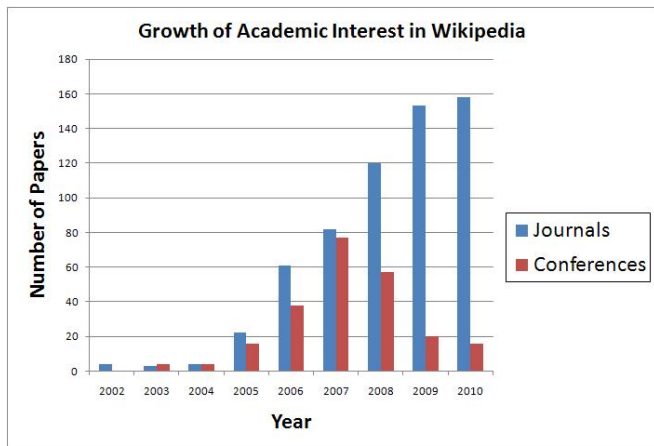


Figure: Comportamento do número de citações.

Conteúdo

Os trabalhos nessa área buscam entender e mensurar a qualidade das informações presentes na enciclopédia.

Pesquisadores da *University of Minnesota*, realizaram um trabalho relevante sobre a diversidade do conteúdo.

[...] Frequent editors dominate what people see when they visit Wikipedia and [...] this domination is increasing.

Mostrar gráfico(s) do artigo

Conteúdo

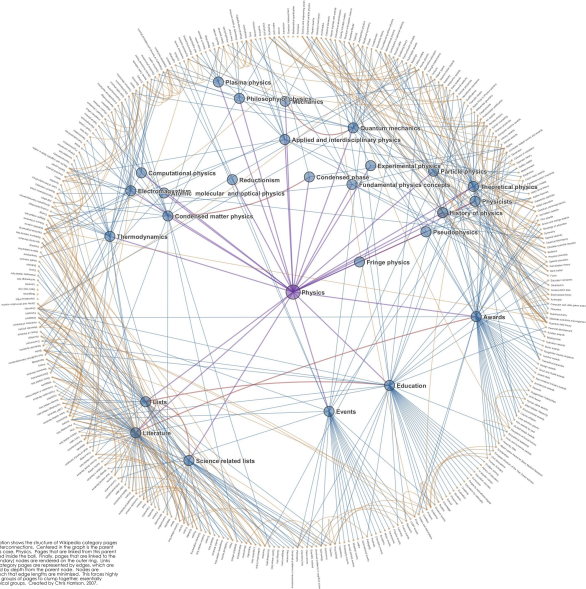
Um [referenciar isso] estudo, conduzido pelo *Oxford Internet Institute*, mostrou desigualdade no número e na qualidade de artigos sobre regiões da Terra. A maior parte das informações diz respeito a regiões da América do Norte, Europa e oeste da Ásia.[3]
Colocar uma imagem aqui! Pode ser do WikiTrip.

Conteúdo

Segundo pesquisadores do *Ubiquitous Knowledge Processing Lab*, a estrutura de categoria da Wikipédia é:

a taxonomy able to compete with WordNet on linguistic processing tasks.

Conteúdo



This visualization shows the structure of Wikipedia category pages and their interconnections. Centred in the graph is the parent node in this case, Physics. Major hub nodes (nodes that are connected to the most other nodes) are highlighted in the outer ring. Lines between category pages are numbered by edges, which are color coded by length from the center node. Nodes are clustered such that edges are color consistent. The far right nodes are grouped into clusters forming topical groups. Checked by Chris Harrison, 2007.

Aspectos sociais

O tamanho do conteúdo e número de usuários faz da Wikipédia um ambiente complexo, livre para estudo de sua estrutura social, e de como ela evolui.

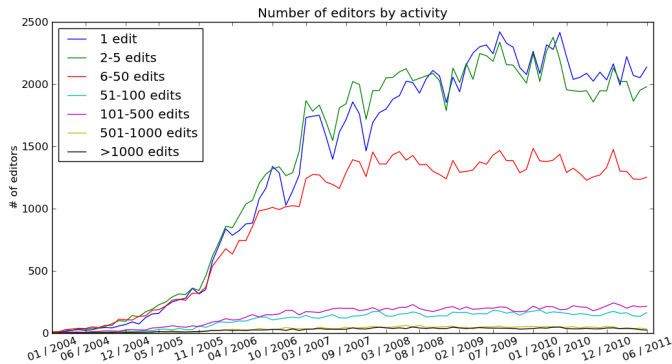


Figure: Segmentos de atividade dos editores ao longo do tempo.[2]

Tendências sociais

A internet pode ser considerada em algumas regiões como um espelho - ainda imperfeito - de sua sociedade real, com seus eventos, complexidade e comportamento.

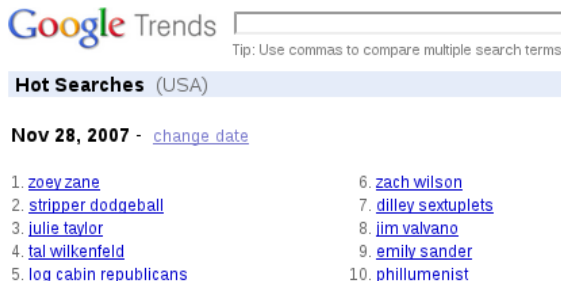


Figure: Lista dos 10 assuntos mais buscados no Google em 28/11/2007 [4]

Tendências sociais

Essa lista devia fazer sentido na época, hoje nem tanto. Então como filtrar a intenção dos leitores?

Tendências sociais

Um teste feito avalia a relação entre o número de casos de dengue no Brasil, nos anos de 2009 e 2010, com os números de acessos às páginas da Wikipédia sobre dengue, e o mosquito *Aedes aegypti*.

Tendências sociais

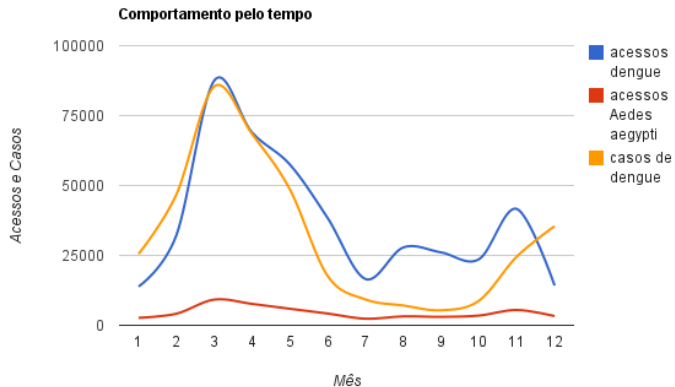


Figure: Para o ano de 2009.

Tendências sociais

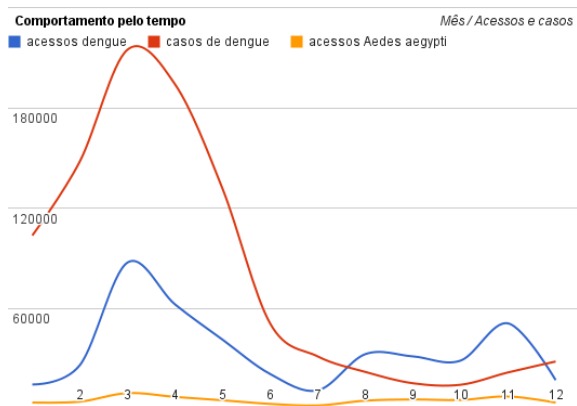


Figure: Para o ano de 2010.

Perguntas?

Referências

-  ClusterBall Visualizing Wikipedia - Chris Harris.
<http://www.chrisharrison.net/projects/clusterball/>
-  Portuguese Wikipedia trends and behavior/Editor Trends -
Wikimedia Foundation, 2011. http://meta.wikimedia.org/wiki/Research:Portuguese_Wikipedia_trends_and_behavior/Editor_Trends
-  Mapping the Geographies of Wikipedia Content - Andrew
Blum, 2009. <http://www.zerogeography.net/2009/11/mapping-geographies-of-wikipedia.html>
-  Google Trends. <http://www.google.com/trends/hottrends?sa=X&date=2007-11-28>