Lecture 22: CS 5306 / INFO 5306: Crowdsourcing and Human Computation

## Course Projects: Amazon Mechanical Turk

- Payments:
  - Put \$20 on account
  - Get reimbursed by Information Science
  - If you need more \$ contact me
- Try simple task
  - (Should have requester account)

# Types of Crowdsourcing

- Overt
  - Collecting (Amazon Reviews)
  - Labor Markets (Amazon Mechanical Turk)
  - Collaborative Decisions (Prediction Markets)
  - Collaborative Creation (Wikipedia)
  - Smartest in the Crowd (Contests)
  - Games with a Purpose
- Covert / Crowd Mining
  - Web page linkage, search logs, social media, collaborative filtering
- Dark side of crowdsourcing and human computation
- Collective intelligence in animals

# Types of Crowdsourcing

- Overt
  - Collecting (Amazon Reviews)
  - Labor Markets (Amazon Mechanical Turk)
  - Collaborative Decisions (Prediction Markets)
  - Collaborative Creation (Wikipedia)
  - Smartest in the Crowd (Contests)
  - Games with a Purpose
- Covert / Crowd Mining
  - Web page linkage, search logs, social media, collaborative filtering
- Dark side of crowdsourcing and human computation
- Collective intelligence in animals

### Mammon and the Archer

• mammon: "material wealth or possessions especially as having a debasing influence" (Mirriam Webster Dictionary)

### Mammon and the Archer

• mammon: "material wealth or possessions especially as having a debasing influence" (Mirriam Webster Dictionary)

archer?

# Mammon and the Archer

Why?

- 1961 Rose Bowl
- Washington Huskies vs Minnesota Golden Gophers
- Half-time show: Card stunt



- 1961 Rose Bowl
- Washington Huskies vs Minnesota Golden Gophers
- Half-time show: Card stunt

- 1961 Rose Bowl
- Washington Huskies vs Minnesota Golden Gophers
- Half-time show: Card stunt
- Image 12: Husky -> Beaver

- 1961 Rose Bowl
- Washington Huskies vs Minnesota Golden Gophers
- Half-time show: Card stunt
- Image 12: Husky -> Beaver
- Image 13: Washington appeared in reverse

- 1961 Rose Bowl
- Washington Huskies vs Minnesota Golden Gophers
- Half-time show: Card stunt
- Image 12: Husky -> Beaver
- Image 13: Washington appeared in reverse
- Image 14









- Worker experience
- Undermining crowds
- Using crowds for nefarious purposes
- Crowd failures

- Worker experience
- Undermining crowds
- Using crowds for nefarious purposes
- Crowd failures

How would you protect against these?

- Worker experience
- Undermining crowds
- Using crowds for nefarious purposes
- Crowd failures

# Worker Experience

- Tuesday's lecture by Brian McInnis
- "Anda's Game" by Cory Doctorow http://www.salon.com/2004/11/15/andas\_game/

# They play games for 10 hours - and earn £2.80 in a 'virtual sweatshop'

Tony Thompson in Caracal, Romania The Observer, Sunday 13 March 2005 01.49 GMT Article history

Bogdan Ghirda is paid £70 a month to do what most bosses would fire him for. From the moment he arrives at work he plays computer games on the internet.

With only a few short breaks Ghirda, 20, goes on playing furiously for 10 hours in the backroom of a run-down apartment block in Caracal, Romania. The moment he leaves his desk a member of an evening shift takes over the computer and continues the same game with equal determination.

Between them, the company's 11 employees keep a dozen or so computers running 24 hours a day, seven days a week.

Although Ghirda works in Romania, the computers and the internet connection he uses are paid for by a company in northern California. Gamersloot.net is one of a growing number of firms taking advantage of a boom in online computer games by opening 'virtual sweatshops', using the low pay in poor countries to provide services for wealthy western players.

Older computer games pit a single player against computerised opponents, but the new ones allow players to join forces with others anywhere in the world. There are now an estimated 350 such games, the most popular of which have more than 300,000 subscribers, each paying a monthly fee to keep their place.

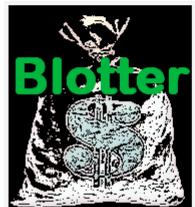
- Worker experience
- Undermining crowds
- Using crowds for nefarious purposes
- Crowd failures

- Worker experience
- Undermining crowds
- Using crowds for nefarious purposes
- Crowd failures

### Nilton Rossoni Sentenced To 68 Months In Federal Prison For Colossal eBay Fraud; Elaborate Scheme Featured 59 Mail Drops, 260 Bogus Auction Accounts

By PatrickPretty.com 6:46 pm Feb 5, 2010





It was a case that was all about the numbers. In the end, the number with the most meaning to Nilton Rossoni was 68 — the number of months he'll be spending in federal prison.

Rossoni conducted more than 5,500 fraudulent auction on eBay. He pulled off his scheme by using at least 260 bogus accounts, at least 59 mail drops, six names, fou bogus passports and three banks.

Rossoni, 50, formerly of Sunny Isles and Hallandale Beach, Fla., collected \$717,000 in the scheme between October 2003 and June 2008. The bizarre fraud was smashed by the U.S. Postal Inspection Service

Winning bidders were notified via e-mail to send a check or money order payable either to Celso Ferreira, Jorge Carlos, Joao Santos, Lourival Philipps, Prime Hill Inc. or Primo Hill Inc. Buyers were instructed to send payments via U.S. Mail to specific addresses, all of which proved to be mail drops.

2 May 2016 25

<sup>&</sup>quot;Elaborate" barely described the scheme.

<sup>&</sup>quot;Rossoni rented at least 59 separate private mail boxes at various Commercial Mail Receiving Agencies (CMRA), including The UPS Store, Mail Boxes Etc., and Pak-Mail, using fraudulent Brazilian passports in the names of Celso Ferreira. Jorge Carlos.

### The Sybil Attack

John R. Douceur Microsoft Research johndo@microsoft.com

"One can have, some claim, as many electronic personas as one has time and energy to create."

- Judith S. Donath [12]

Abstract – Large-scale peer-to-peer systems face security threats from faulty or hostile remote computing elements. To resist these threats, many such systems employ redundancy. However, if a single faulty entity can present multiple identities, it can control a substantial fraction of the system, thereby undermining this redundancy. One approach to preventing these "Sybil attacks" is to have a trusted agency certify identities. This paper shows that, without a logically centralized authority, Sybil attacks are always possible except under extreme and unrealistic assumptions of resource parity and coordination among entities.

### 1. Introduction

We argue that it is practically impossible, in a distributed computing environment, for initially unknown remote computing elements to present convincingly distinct identities. With no logically central, trusted authority to vouch for a one-to-one correspondence between entity and identity, it is always possible for an unfamiliar entity to present

If the local entity has no direct physical knowledge of remote entities, it perceives them only as informational abstractions that we call *identities*. The system must ensure that distinct identities refer to distinct entities; otherwise, when the local entity selects a subset of identities to redundantly perform a remote operation, it can be duped into selecting a single remote entity multiple times, thereby defeating the redundancy. We term the forging of multiple identities a *Sybil attack* [30] on the system.

It is tempting to envision a system in which established identities vouch for other identities, so that an entity can accept new identities by trusting the collective assurance of multiple (presumably independent) signatories, analogous to the PGP web of trust [37] for human entities. However, our results show that, in the absence of a trusted identification authority (or unrealistic assumptions about the resources available to an attacker), a Sybil attack can severely compromise the initial generation of identities, thereby undermining the

2 May 2016 26

### Shilling Recommender Systems for Fun and Profit

Shyong (Tony) K. Lam

John Riedl

GroupLens Research Computer Science and Engineering University of Minnesota Minneapolis, MN 55455

{lam, riedl}@cs.umn.edu

### ABSTRACT

Recommender systems have emerged in the past several years as an effective way to help people cope with the problem of information overload. One application in which they have become particularly common is in e-commerce, where recommendation of items can often help a customer find what she is interested in and, therefore can help drive sales. Unscrupulous producers in the never-ending quest for market penetration may find it profitable to shill recommender systems by lying to the systems in order to have their products recommended more often than those of their competitors. This paper explores four open questions that may affect the effectiveness of such shilling attacks: which recommender algorithm is being used, whether the application is producing recommendations or predictions, how detectable the attacks are by the operator of the system, and what the properties are of the items being attacked. The questions are explored experimentally on a large data set of movie ratings. Taken together, the results of the paper suggest that new ways must be used to evaluate and detect shilling attacks on recommender systems

recommendations to a user regarding which items she may find interesting. One instance of a recommender system is *MovieLens* (http://www.movielens.org). GroupLens, our research group, operates this recommender system, which makes personalized recommendations suggesting movies that a user might like based on movies that she has seen and has expressed an opinion about.

While recommender systems are clearly beneficial to users, they can also be a valuable asset to retail companies in helping their customers find things that they might want to buy and, in effect, increasing not only sales, but perhaps also cross-sales and customer retention. This is particularly true in the realm of e-commerce. For example, *Amazon.com* has made many recommender systems available to their customers. These range from manually operated recommenders where users can recommend items to other users by writing reviews or creating lists, to automated systems where the site generates a list of recommended items based on what the user has looked at recently or has purchased in the past.

Producers of items (manufacturers, authors, etc.) would like their products to sell well in the marketplace. With recommender

### Wherefore Art Thou R3579X? Anonymized Social Networks, Hidden Patterns, and Structural Steganography

Lars Backstrom
Dept. of Computer Science
Cornell University, Ithaca NY
lars@cs.cornell.edu

Cynthia Dwork Microsoft Research dwork@microsoft.com Jon Kleinberg
Dept. of Computer Science
Cornell University, Ithaca NY
kleinber@cs.cornell.edu

#### ABSTRACT

In a social network, nodes correspond to people or other social entities, and edges correspond to social links between them. In an effort to preserve privacy, the practice of anonymization replaces names with meaningless unique identifiers. We describe a family of attacks such that even from a single anonymized copy of a social network, it is possible for an adversary to learn whether edges exist or not between specific targeted pairs of nodes.

### Categories and Subject Descriptors

F.2.2 [Analysis of Algorithms and Problem Complexity]: Nonnumerical Algorithms and Problems

### **General Terms**

Theory, Measurement

### Keywords

social networks, anonymization, privacy in data mining

#### . INTRODUCTION

be considering the "purest" form of social network data, in which there are simply nodes corresponding to individuals and edges indicating social interaction, without any further annotation such as time-stamps or textual data.

In designing studies of such systems, one needs to set up the data to protect the privacy of individual users while preserving the global network properties. This is typically done through anonymization, a simple procedure in which each individual's "name" e.g., e-mail address, phone number, or actual name - is replaced by a random user ID, but the connections between the (now anonymized) people - encoding who spoke together on the phone, who corresponded with whom, or who instant-messaged whom - are revealed. The motivation behind anonymizing is roughly as follows: while the social network labeled with actual names is sensitive and cannot be released, there may be considerable value in allowing researchers to study its structure. For such studies, including those cited above, researchers are not specifically interested in "who" corresponds to each node, but in the properties of the graph, such as its connectivity, node-to-node distances, frequencies of small subgraphs, or the extent to which it can be clustered. Anonymization is thus intended to exactly preserve the pure unannotated structure of the graph while suppressing the "who" information.

Can this work? The hope is that being handed an anonymized picture of a social network — just a graph with a random identifier attached to each node — is roughly akin to being given the

2 May 2016 28

# Google Let JC Penney Spam Search Results For Months

Matt Rosoff | Feb. 13, 2011, 12:10 PM | ♠ 5,901 | ■ 15

The New York Times exposed the dirty side of search engine optimization this morning with a long article about how JC Penney spammed Google so it would appear at the top of search results.

Somebody created thousands of

Somebody created thousands of fake pages with the keywords that Penney wanted to game, like "black dresses," and a direct link to Penney's site. This messes with Google's PageRank algorithm, which assumes that a site is useful if it's popular. (A Penney spokesperson denied that the company knew what was going on - it was probably a guerrilla SEO team or agency working on Penney's behalf.)

The amazing part of the story isn't how Penney tricked Google – this kind of "black hat" SEO has been around almost since Google began.

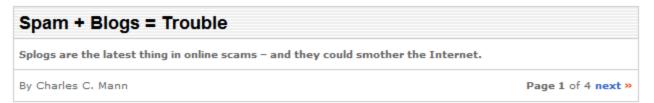


Google cofounders Larry Page and Sergey Brin back in more carefree days.



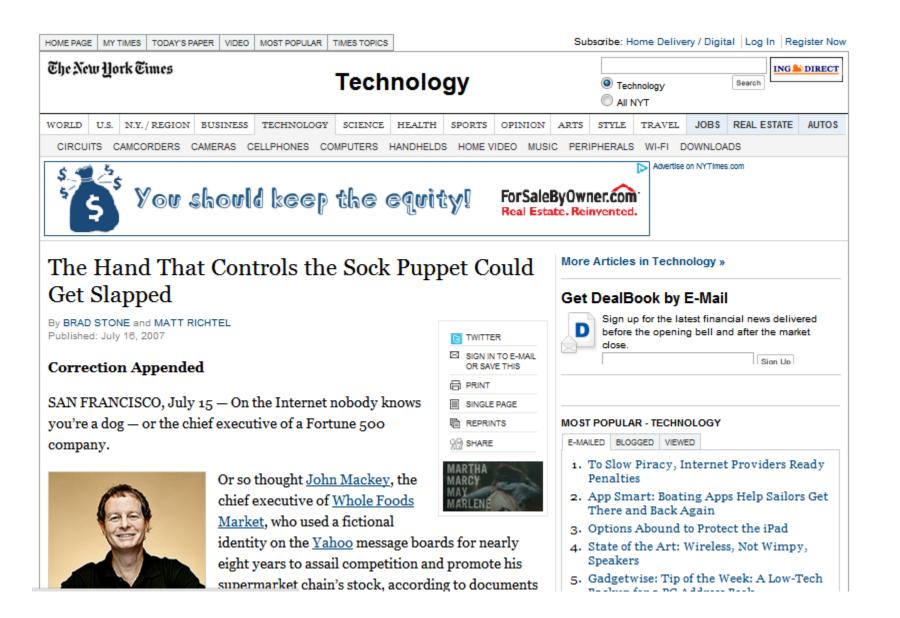
Issue 14.09 - September 2006

Subscribe to WIRED magazine and receive a FREE gift!



I am aware that spending a lot of time Googling yourself is kind of narcissistic, OK? But there are situations, I would argue, when it is efficiently – even forgivably – narcissistic. When I published a book last year, I wanted to know what, if anything, people were saying about it. Ego-surfing was the obvious way to do that. Which is how I stumbled across Some Title.





### google.org Flu Trends

### Google.org home

Flu Trends

Home

Select country/region

•

How does this work?

FAQ

Flu activity

Intense

High

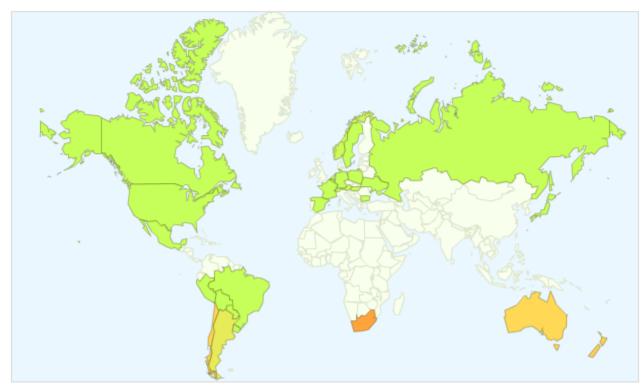
Moderate

Low

Minimal

### Explore flu trends around the world

We've found that certain search terms are good indicators of flu activity. Google Flu Trends uses aggregated Google search data to estimate flu activity. Learn more »



<u>Download world flu activity data</u> - <u>Animated flu trends for Google Earth</u> - <u>Compare flu trends across regions in Public Data Explorer</u>

On [[March 24]] [[1989]], shortly after midnight, the oil tanker [[Exxon Valdez]] struck [[Bligh Reef]] in [[Prince William Sound]], [[Alaska]], spilling more than 11 million gallons (42,000 m&sup3) of crude oil. The spill was the largest in U.S. history, and in the aftermath of the Exxon Valdez incident [[Congress of the United States|U.S. Congress]] passed the [[Oil Pollution Act of 1990]]. Exxon Mobil has not yet paid the \$5 billion in spill damages it owes to the 32,000 Alaskan fishermen.

On [[March 24]] [[1989]], shortly after midnight, the oil tanker [[Exxon Valdez]] struck [[Bligh Reef]] in [[Prince William Sound]], [[Alaska]], spilling more than 11 million gallons (42,000 m&sup3) of crude oil. The spill was the largest in U.S. history, and in the aftermath of the Exxon Valdez incident [[Congress of the United States|U.S. Congress]] passed the [[Oil Pollution Act of 1990]]. ExxonMobil paid \$300 million immediately and voluntarily to more than 11,000 Alaskans and businesses affected by the Valdez spill. In addition, the company + paid \$2.2 billion on the cleanup of Prince William Sound, staying

with the cleanup from 1989 to 1992, when the State of Alaska and the U.S. Coast Guard declared the cleanup complete.

ExxonMobil also has paid \$1 billion in settlements with the state and federal governments. Virtually all Valdez compensatory damages were paid in full within one year of the accident, and the trial court commended ExxonMobil for coming forward "with its people and its pocketbook and doing what had to be done under difficult circumstances."

# WikiScanner

- Virgil Griffith
- 2007-2013?

In [[August 2003]], [[Penguin Books]] published Franken's "[[Lies and the Lying Liars Who Tell Them|Lies and the Lying Liars Who Tell Them: A Fair and Balanced Look at the Right]]". [[Fox News]] sued, claiming that Franken [[infringement|infringed]] its registered [[trademark]] rights in the phrase, "Fair and Balanced." Fox was unsuccessful, with a federal judge finding the [[lawsuit]] to be "wholly without merit." The lawsuit focused a great deal of media attention upon Franken's book and greatly enhanced its sales. Reflecting later on the lawsuit during an interview on the [[National Public Radio]] program "[[Fresh Air]]" on [[September 3]], [[2003]], Franken said that Fox's case against him was "literally laughed out of court" and that "wholly (holy) without merit" is a good characterization of Fox News itself.

In [[August 2003]], [[Penguin Books]] published Franken's "[[Lies and the Lying Liars Who Tell Them|Lies and the Lying Liars Who Tell Them: A Fair and Balanced Look at the Right]]". [[Fox News]] sued, claiming that Franken [[infringement|infringed]] its registered [[trademark]] rights in the phrase, "Fair and Balanced." Fox was unsuccessful, with a federal judge finding the [[lawsuit]] to be "wholly

+ without merit." The lawsuit focused a great deal of media attention upon Franken's book and greatly enhanced its sales. Reflecting later on the lawsuit during an interview on the liberal [[National Public Radio]] program "[[Fresh Air]]" on [[September 3]], [[2003]], Franken said that Fox's case against him was the best thing to happen to his book sales.

2 May 2016 35

Many feel that copyleft licenses are desirable and popular for shared works precisely because they "are" viral, and apply to all derivative works, which are thus "infected" by the requirement to re-integrate changes deemed desirable by any party down the line. This requirement is seen as important because it ensures uniform license terms and free access, and makes copyleft projects resistant to unnecessary [[fork (software)|forking]] because all maintainers of the original work, or other versions, - may use any modifications released by anyone. Useful changes tend to be merged, and different versions are maintained only to the extent that they are useful. Without the "viral" license, variant terms can apply to the forks and derivative works can be controlled commercially by the parties that extend or translate them, which can be considered as some of the disadvantages of non-copyleft "open source" projects. It is thought that [[Linux]] has not suffered the same [[Unix wars|fragmentation]] as [[Unix]] because it is copylefted.

Opponents of copyleft point out that even using a single line of copy-lefted code inside millions of line of non-copylefted code will cause all the code to automatically get copylefted. This is not unlike the behaviour of a computer virus or a biological virus which infects a larger entity though it's own dimensions are small in comparison.

2 May 2016 36

Since the late [[1970]]s, Nestlé has attracted much criticism for its baby milk marketing policies in [[Developing country|developing countries]]. This has centered on its apparent recommendations for [[Breastfeeding|nursing]] mothers to switch to its [[infant formula]] milk products, leading to the alleged deaths of about 1.5 million babies each year as a result of formula being mixed with contaminated water{{fact}}. Nestlé allegedly has violated the widely agreed-upon [[International Code of Marketing of Breast-milk Substitutes]]{{fact}}. This has led to a boycott coordinated by the International Nestlé Boycott Committee, informed by monitoring conducted by the [[International Baby Food Action Network]]. In 1982, Nestlé implemented the [[WHO]] Code of Marketing of Breast-Milk Substitutes in developing countries. The instructions were reviewed and refined in 1984 in consultation with the WHO, [[UNICEF]] and the International Nestlé Boycott Committee.

Nestlé itself still advertises breast milk replacements, and does use pictures of babies in its advertising{{citation needed}}. According to the report "Breaking the Rules, Stretching the Rules 2004" put out by IBFAN/ICDC it is still continuing aggressive and - misleading advertising tactics

[http://www.ibfan.org/english/codewatch/btr04/btr04contents.html]. Nestlé has issued instructions to all its offices to ensure strict compliance with the International Code. However, the boycott still continues in some 20 countries.

In 1982, Nestlé implemented the [[WHO]] Code of Marketing of Breast-Milk Substitutes in developing countries. The instructions were reviewed and refined in 1984 in consultation with the WHO, [[UNICEF]] and the International Nestlé Boycott Committee.

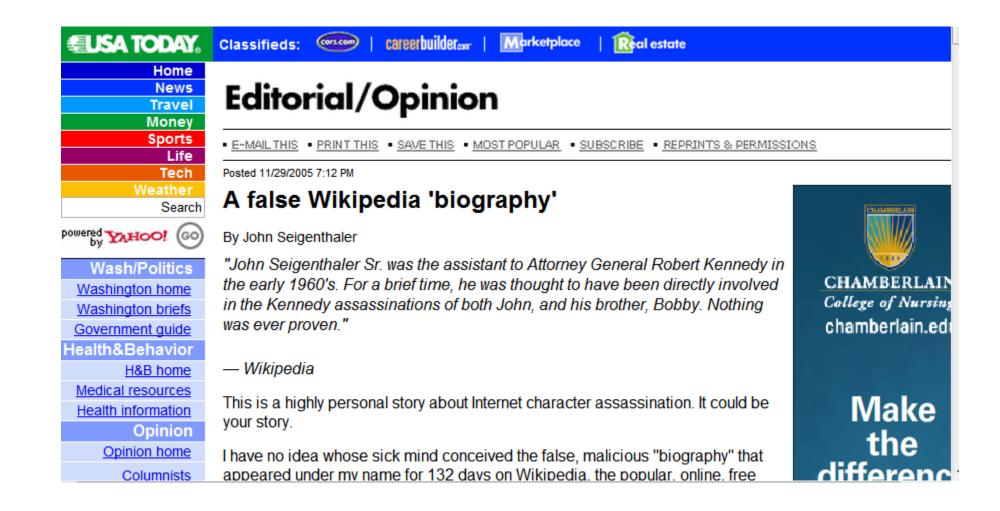
- ==Environmental and human rights controversies==
- ===Union Carbide Bhopal disaster===
- {{main article|Bhopal disaster}}

In 1984, a chemical factory operated by [[Union Carbide]], an American company, leaked lethal gases into the surrounding environment, which has caused more than 20,000 deaths and 100,000 disabilities. In 1999, Dow purchased Union Carbide, which is now a [[wholly owned subsidiary]]. Dow denies legal liability for the disaster, since it did not own or operate the Bhopal factory. <ref>{(cite website | -url=http://news.bba.co.uk/onthisday/hi/dates/stories/december/3/newsid\_2698000/2698709.stm | title=On This Day: December 3, 1984 | publisher=British Broadcasting Corporation | accessed=June 24, 2006}} </re>

The human rights organization [[Amnesty International]] has held Dow responsible for the persistence of contamination and continued suffering of the victims of Bhopal disaster.<ref>{[cite website | url=http://web.amnesty.org/library/pdf/ASA200152004ENGLISH/\$File/ASA2001504.pdf | title=Clouds of Injustice: Bhopal Disaster, 20 years on | date=2004 | publisher=Amnesty InternationI]}</ri>
Victims of the Bhopal disaster and their supporters have demanded that Dow compel Union Carbide to face trial in India, where it is a fugitive from justice. <ref> [http://www.studentsforbhopal.org/Assets/CJM\_Jan2005.pdf]</ref>

- === Breast implants ===

A major manufacturer of silicone [[breast implants]], Dow Corning was successfully sued in 1977 for damages arising from a woman whose implants ruptured; it was the first such successful suit, and Dow Corning paid \$170,000 in a settlement. During the 1980s, [[Ralph Nader]]'s [[Public Citizen]] Health Research Group publicised its belief that the implants were cancer-causing in December of 1990, and





Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction

Help

About Wikipedia Community portal Recent changes Contact page

#### Tools

What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item

Print/export

Create a book

Download on DDE

Project page Talk Read View source View history Search Q

### Wikipedia:Biographies of living persons

From Wikipedia, the free encyclopedia



If you are concerned about the accuracy or appropriateness of biographical material on Wikipedia, report problems at the biographies of living persons noticeboard. For articles about yourself, please see "Dealing with articles about yourself" below.

Shortcuts: WP:ALIVE WP:BLP WP:LIVE



This page documents an English Wikipedia policy, a widely accepted standard that all editors should normally follow. Changes made to it should reflect consensus.



This page in a nutshell: Material about living persons added to *any* Wikipedia page must be written with the greatest care and attention to verifiability, neutrality, and avoidance of original research.

Find this page confusing? Just use this link to ask for help on your talk page; a volunteer will visit you there shortly!

Editors must take particular care when adding **information about living persons** to *any* Wikipedia page.<sup>[1]</sup> Such material requires a high degree of sensitivity, and must adhere *strictly* to all applicable laws in the United States, to this policy, and to Wikipedia's three core content policies:

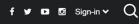
- Neutral point of view (NPOV)
- Verifiability (V)
- No original research (NOR)

We must get the article *right*. Be very firm about the use of high-quality sources. All quotations and any material **challenged or likely to be challenged** must be explicitly attributed to a reliable, published source, which is usually done with an inline citation. Contentious material about living persons (or, in some cases, recently deceased) that is unsourced or poorly sourced – whether the material is negative, positive, neutral, or just questionable – should be **removed immediately and without waiting for discussion**.<sup>[2]</sup> Users who persistently or egregiously violate this policy may be blocked from editing





### **CULTURE**





Assistant,
Rapid Respon...
Syracuse, NY

► Utility Person, Keith Titus ... Weedsport, NY ► Bioanalytical LCMS Pro..., Quintiles Ithaca, NY

Search Jobs >>

# Wikipedia's longest hoax ever gets busted after more than ten years

Kim Renfro and Gus Lubin

⊙ Feb. 4, 2016, 10:31 AM 

◆53,406 

○ 1















### **RECOMMENDED FOR YOU**



Sunday's 'Game of Thrones' had a clever Easter egg you probably missed

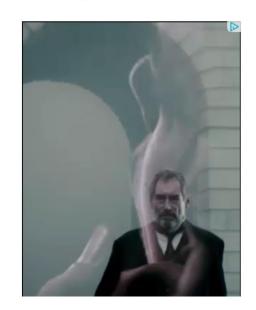


This scene from the 'Game of Thrones' premiere is much more emotional when you realize its significance

Wikipedia editor Calamondin12 knew this looked bad.

The article about "Jack Robichaux" had been flagged as an orphan, meaning it had few or no inbound links. What's more, the article featured a racist comment, and the only source provided did not mention anyone by that name. Worst of all, the article had been on Wikipedia for more than ten years.





& Log in / create account

Q



Main page Contents Featured content Current events Random article Donate to Wikipedia

- Interaction
   Help
   About Wikipedia
   Community portal
   Recent changes
   Contact Wikipedia
- ▼ Toolbox
  What links here
  Related changes
  Upload file
  Special pages
  Permanent link

Article Discussion Read Edit View history Search

### **Grass Mud Horse**

From Wikipedia, the free encyclopedia

The **Grass Mud Horse** or **Cao Ní Ma** is a Chinese Internet meme widely used as symbolic defiance of the widespread Internet censorship in China. It is one of the 10 mythical creatures, and since an article about it was created on Baidu Baike in early 2009, it has become a cult phenomenon on the internet in China through chat forums. Videos, cartoons and merchandise of this animal, which apparently resembles the alpaca, have appeared, and it has since received worldwide press attention.

#### Contents [hide]

- 1 Etymology and species
- 2 Habitat
- 3 Formats
- 4 Censorship
- 5 See also
- 6 References
- 7 External links

### Etymology and species

Cao Ni Ma, literally "Grass Mud Horse", was supposedly a species of alpaca. The name is derived from a profanity



A Grass Mud Horse plush doll Traditional Chinese 草泥馬

[edit]

2 May 2016 42

- Worker experience
- Undermining crowds
- Using crowds for nefarious purposes
- Crowd failures

- Worker experience
- Undermining crowds
- Using crowds for nefarious purposes
- Crowd failures