Wikipedia and the Problem of Authorship. Aaron Swartz’s Hypothesis

Krzysztof Gajewski
Institute of Literary Research
Polish Academy of Science

Oradea/Nagyvárad, 7-8 October 2016
The Gang of 500 vs. The Anonymous Horde

Alternatives: Persistent Word View and Persistent Word Revision

Verification of Swartz’s thesis and Problems Arising
  The Double Face of Wiki Text
  Concise as a Wikipedia Article
  The Authors
  Copied Content Problem

Conclusions. Ritual vs. Conflict View of Communication
English Wikipedia Size in Volumes

- 12 stacks
- 2319 volumes (Britannica size)
- 5.2 million articles (as of July 2016)
- 3.068 billion words (as of July 2016)
“you cannot change the content of Wikipedia, you can only supply a new one. ”
Revision history of a Wikipedia article
Model of Communication

- Receiver = we
- Sender = ?

„the free encyclopedia that anyone can edit.”
The Gang of 500 vs. The Anonymous Horde
Jimmy Wales:

- 73.4% of all the edits — 2% of the users (1400 people)
- 50% of all the edits — 0.7% of the users (524 people)$^1$

$^1$Swartz 2006
What Is an Edit?

Two types of edits:

1. uploading textual content
2. „wikisation” of a content already uploaded
Top 10 contributors to „Alan Alda” entry

1. by edits: 7 registered vs. 3 anonymous
2. by letters added: 8 anonymous vs. 2 registered
Apparent exceptions

1. translations
2. plagiarism
Alternatives: Persistent Word View and Persistent Word Revision
Persistent Word View (PWV) is based on

1. number of letters she input
2. the popularity of the content

Top 10% most active editors generated 86% of Persistent Word View (Feb. 2006)

\[2\] Priedhorsky 2007
Persistent Word Revision (PWR) is

*The sum total of subsequent revisions persisted by the words in a revision.*\(^3\)

---

\(^3\)Research:Content persistence
• PWV stresses a role of a reader
• PWR — editors decide of the value of the content
Verification of Swartz’s thesis and Problems Arising
30 random entries from Polish Wikipedia (“Random article”):

1. Powstanie Kantonalistów
2. Gmina Czarnylas
3. Aurora (telenowela)
4. Podróż na Tajemniczą Wyspę
5. Oberwiera
6. Park Narodowy Altaj-Tawanbogd
7. Kościół św. Mikołaja w Wilnie
8. Klafer am Hochficht
9. Anthaxia attenuata
10. Jerzy Panek (polityk)
11. Wojna z terroryzmem
12. FSO Polonez Analog
13. Strange Frontier
14. Praszywe (Dolina Łatana)
15. Chelmsley Wood
16. Droga krajowa B5 (Niemcy)
17. San Roque (Mariany Północne)
18. Krupy (powiat sokołowski)
19. Alex MacDowall
21. Bitwa pod Olszanicą
22. Rodrigo Oliveira de Bittencourt
23. Hrabstwo White (Tennessee)
24. Bradford (Ohio)
25. Most Królowej Jadwigi w Poznaniu
26. Walenty Forys
27. (Get A) Grip (On Yourself)
28. Andrzej Ekiert
29. Droga krajowa nr 471 (Węgry)
30. Wyganki
The Double Face of Wiki Text
Thunder (mascot)

From Wikipedia, the free encyclopedia

Thunder is the stage name for the horse who is the official live animal mascot for the Denver Broncos football team. Three purebred horses whose coats lightened with age until they turned completely white. Sharon Magness-Blake has owned all three horses, an trainer since 1999. As of 2016, Thunder has appeared in four Super Bowls with the team since 1998. The original Thunder performed Thunder III appeared in Super Bowl XLVIII and Super Bowl 50. Thunder III also made appearances in Times Square and on television the pre-game promotion for Super Bowl XLVII. He routinely appears in parades, makes hospital and school visits, and attends various events, ridden in elevators, and appeared indoors at press conferences and banquets.

Thunder's duties as mascot typically include leading the team onto the field at the start of every home game, and a gallop down the touchdow or field goal. Thunder and his rider also interact with fans before the game; the horse is particularly popular with children served as Thunder need to remain calm in situations that would normally frighten most horses, such as being in football stadiums pyrotechnics, cheerleaders waving pom-poms, and other spectacles common to National Football League (NFL) games. Thunder a horse head mask atop a Broncos uniform.

The original Thunder, later named "Thunder, Sr.", was described as bold and courageous. He was a stallion registered as JE Kob the Broncos from 1993 until his retirement in 2004. He continued making community appearances until his death in 2009. Thunde Arabian gelding registered as Winter Solstice. He had been the personal pleasure riding horse of Magness-Blake. Judge describ his roe as mascot but eventually grew into it. He retired from mascot duties in early 2014 but appeared in the 2016 Super Bowl 51 Broncos win over the Carolina Panthers. "Thunder II", a gelding registered as Me N Myshadow, was the understudy to Thunder II, at age three when he was started under saddle. He began performing at preseason games in 2013. Although Thunder II was still Thunder III was sent to Super Bowl XLVII because he was younger and better able to handle air travel. He is described as laid-back performing.
"Thunder" is the [[stage name]] for the [[horse]] who is the official live animal [[mascot]] for the [[Denver Broncos]]. [[Arabian horse|Arabians]] have held this role since 1993, all [[gray horse|gray horses]] whose coats lighten with age (a trait known as [[blaze]], [[Harbor Magnes-Blake]] has owned all three horses, and Ann Judge has been their rider since 1998 and trainer since 1998. The original Thunder performed in [[Super Bowl XXXII]] and [[Super Bowl XXXIII]] and [[Super Bowl XXXIV]]. Thunder III also made appearances in [[Time's Square]] and on television morning news shows in [[New York City]]. He routinely appears in parades, makes hospital and school visits, and attends various other public functions and events. Thunder's duties as mascot typically include leading the team onto the field at the start of every [[home game field]] whenever the team scores a [[touchdown]] or [[Field goal (American and Canadian football)|field goal]]. Thunder and the horse are particularly popular with children, who are allowed to pet him. The horses who have served as Thunder need to be gentle, calm, and easy to handle, with a temperament that makes them suitable for parades and public appearances.

The original Thunder, later named "Thunder, Sr.", was described as bold and courageous. He was a [[stallion]] [[gray horse]] who was known for his speed and agility. Thunder, Jr. was retired in 2004 due to an injury sustained during a practice session. Thunder, III, a [[gray horse]], was acquired in 2004 and became the [[understudy]] to Thunder II, and trained specifically for mascot duties beginning at age three when he was performing at preseason games in 2013. Although Thunder II was still active as team mascot during the 2013-14 season, Thunder III was chosen as the new mascot due to his younger and better ability to handle air travel. He is described as laid-back, preferring to doze off during games when not in use.

**History**

Video clip of Thunder III (Ne M Nyshadow) making a run in 2014 after [[Pittsburgh Steelers]] touchdown (2016).
IVA vs. OVA

- IVA is easy to determine, calculated automatically, but includes WikiMarkUp
- OVA is what we actually see, what counts for a reader, but is difficult to precise
Methodological problems

- Wiki MarkuUp Language
- Non-textual content: links, bibliography, tables, sounds, images, graphics, diagrams, schemes etc.
- Styles: headings, lists, tables etc.
- Tranclusion (templates)
- Photos and other multimedia (WikiCommons)
Concise as a Wikipedia Article
The size of articles

<table>
<thead>
<tr>
<th></th>
<th>Size of Articles</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>developed</td>
<td>more than 10 lines</td>
<td>6</td>
<td>20%</td>
</tr>
<tr>
<td>somehow developed</td>
<td>4-10 lines</td>
<td>6</td>
<td>40%</td>
</tr>
<tr>
<td>undeveloped</td>
<td>less than 4 lines</td>
<td>12</td>
<td>40%</td>
</tr>
</tbody>
</table>

- only 13% of articles had more than 50 revisions.
- average OVA in the research sample: 1961 chars
- average IVA in the research sample: 3960 chars
- average IVA in the Polish Wikipedia: 2718 chars (Feb., 2014)
The Authors
Presumptions

- In most of the cases articles — up to 2 authors.
- The First Author — the biggest contribution in terms of OVA
- The Second Author — the second biggest contribution in terms of OVA
The First Author

- 11 of 30 articles — only one human author (9 of 11 — a bot contribution)
- The whole group of the First Authors consists of 29 members.
- A Wikipedist Cojan (ranked as 21st prolific editor in Wikipedia) happened to be the First Author of two random articles.
- 2 of 29 First Authors happened to be a bot: Tsca.bot and MalarzBOT.
- Only 2 of 29 First Authors were anonymous
The Second Author

• 8 of 30 articles — the Second Author added some facts or notes, 22 of 30 — redaction & wikisation only

• 2 of 8 factual intervention was made by anonymous, 6 of 8 by registered users.

• A Wikipedist Lowdown (one of the most prolific editors in Wikipedia) happened to be the Second Author of 2 random articles.
### Authors

The characteristics of 1st and 2nd authors of articles

<table>
<thead>
<tr>
<th></th>
<th>all</th>
<th>registered</th>
<th>anonymous</th>
<th>bot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st authors</td>
<td>29</td>
<td>25 (86%)</td>
<td>2 (7%)</td>
<td>2 (7%)</td>
</tr>
<tr>
<td>2nd authors</td>
<td>21</td>
<td>14 (67%)</td>
<td>4 (19%)</td>
<td>3 (14%)</td>
</tr>
</tbody>
</table>
Copied Content Problem
• At 10:34 am, on 5 August 2005, an anonymous user A added a sentence:

„Such understood „war on terrorism” caused 25 000 civilian deaths”

• At 6:01 pm, on 29 August 2005 another anonymous user B removed this sentence and wrote instead a new one:

„They [opponents of USA policy] also point out the numerous civilian casualties (several thousand) due to military operations.”

• Result?

  • The user B was granted the authorship of the sentence.
  • The user A hasn’t been counted as an author of the final version of the entry at all.
Conclusions. Ritual vs. Conflict View of Communication
### Results

#### OVA Contribution

<table>
<thead>
<tr>
<th></th>
<th>chars</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>registered</td>
<td>40940</td>
<td>90.78%</td>
</tr>
<tr>
<td>anonymous</td>
<td>3353</td>
<td>7.43%</td>
</tr>
<tr>
<td>bots</td>
<td>806</td>
<td>1.79%</td>
</tr>
</tbody>
</table>
Pareto Principle vs. Long Tail

- Gang of 500 Theory $\rightarrow$ Pareto principle
- Anonymous Horde $\rightarrow$ Long Tail (Anderson 2006)
„In a ritual definition, communication is linked to terms such as ’sharing,’ ’participation,’ ’association,’ ’fellowship,’ and ’the possession of a common faith.’ This definition exploits the ancient identity and common roots of the terms ’commonness,’ ’communion,’ ’community,’ and ’communication.’ A ritual view of communication is directed not toward the extension of messages in space, but toward the maintenance of society in time; not the act of imparting information, but the representation of shared beliefs.”

---

4Carey 2009: 15
Conflict View of Communication

ARE YOU COMING TO BED?

I CAN'T. THIS IS IMPORTANT.
WHAT?

SOMEONE IS WRONG ON THE INTERNET.

Conflict View of Communication$^5$

$^5$Conf. Jemielniak 2006: 124


