

# Improving Deliberation and Dispute Resolution on the Web: Investigating Requests for Comment

1. Introduction
2. Quantitative + Qualitative study of RfCs
3. System-building: developing Wikum

*Jane Im*

*This work is ongoing and guided by Prof. David Karger and Ph.D. student Amy X. Zhang.  
The slides contain main contributions made by Jane Im.*

# Introduction

- 1.What are Requests for Comment (RfC)?
- 2.Research Questions
- 3.Assumptions
- 4.Jane's Main Contributions

# 1. What are Requests for Comment (RfC)?

- [Requests for Comment \(RfC\)](#) is a system in Wikipedia for inviting new inputs to help resolve an ongoing content dispute.
- It sometimes requires a closing process in which a neutral editor flags the end of the dispute by writing a closing statement.
- The closing process is sometimes collaborative, but other discussions can remain stale without closures.

RfC that is closed. →  
The consensus is written  
in the closing statement.

## RfC on income inequality effects

*The following discussion is closed. **Please do not modify it.** Subsequent comments should be made on the appropriate discussion page. No further edits should be made to this discussion.*

Consensus was to "omit" the material due to concerns that it is off topic. [Morph](#) <sup>(talk)</sup> 14:31, 17 January 2014 (UTC)

Instead of edit-warring over this excerpt, we clearly need an RFC.

Inequality in land and income ownership is negatively correlated with subsequent economic growth. A strong demand for redistribution will occur in societies where a large section of the population does not have access to the productive resources of the economy. Rational voters must internalize such issues. (Alesina, Alberto (1994). "[Distributive Politics and Economic Growth](#)" <sup>(PDF)</sup>. *Quarterly Journal of Economics*. **109** (2): 465–90. doi:10.2307/2118470. Retrieved 17 October 2013. <sup>Unknown parameter |coauthors= ignored (|author= suggested) (help)</sup>; <sup>Unknown parameter |month= ignored (help)</sup>) High unemployment rates have a significant negative effect when interacting with increases in inequality. Increasing inequality harms growth in countries with high levels of urbanization. High and persistent unemployment also has a negative effect on subsequent long-run economic growth. Unemployment may seriously harm growth because it is a waste of resources, because it generates redistributive pressures and distortions, because it depreciates existing human capital and deters its accumulation, because it drives people to poverty, because it results in liquidity constraints that limit labor mobility, and because it erodes individual self-esteem and promotes social dislocation, unrest and conflict. Policies to control unemployment and reduce its inequality-associated effects can strengthen long-run growth. (Castells-Quintana, David (2012). "[Unemployment and long-run economic growth: The role of income inequality and urbanisation](#)" <sup>(PDF)</sup>. *Investigaciones Regionales*. **12** (24): 153–173. Retrieved 17 October 2013. <sup>Unknown parameter |coauthors= ignored (|author=</sup>

RfC: Images used for Planet Nine

(Notifying previously involved editors: Jehochman, prokaryotes, Serendipodous, Fut.Perf. ♂, Ephraim33, Nergaal, Neutron, Leitmotiv, Kheider, Wnt, Nowa, Itu, Smkolins, Tom Ruen and Jonathunder.)

I really think we need more input regarding the images used on this article. There have been a few previous discussions<sup>[1][2][3][4][5]</sup> but no clear consensus was demonstrated. I propose that we pool our collective opinions here and put things to a vote. Since there are 2 questionable images I'll split this into 2 subsections.

Artist's impression in the infobox

The infobox currently shows an artist's impression of Planet Nine (right), which appears to be closely based on [an image released by Caltech credited to R. Hurt \(IPAC\)](#)<sup>ⓘ</sup>. While artist's impressions may help to grab the reader's attention I do not think that it's becoming of an encyclopedia to reproduce them here and I propose that it be removed, or at the very least removed from the infobox. The only information it conveys are *basic assumptions*, which could easily mislead the reader. [A view of the Earth can be found here](#)<sup>ⓘ</sup> in which the planet's size and distance from the Sun is similarly *not* conveyed. Please bear in mind that even if you don't find the picture misleading, others undoubtedly will.

**Propose** the complete removal of artist's impression. Sorry. nagualdesign 03:43, 4 February 2016 (UTC)

Plenty of artist's impressions are in infoboxes in this encyclopedia. There is no rational reason for its removal. The only issue I have with it is that it is unclear where the Sun is, and how far away it is. Serendipodous 08:07, 4 February 2016 (UTC)

File:Planet-Nine-in-Outer-Space-artistic-depiction.jpg  
An artist's impression of Planet Nine

← Unclosed

RFC on income inequality effects

*The following discussion is closed. **Please do not modify it.** Subsequent comments should be made on the appropriate discussion page. No further edits should be made to this discussion.*

Consensus was to "omit" the material due to concerns that it is off topic. Morphh <sup>(talk)</sup> 14:31, 17 January 2014 (UTC)

Instead of edit-warring over this excerpt, we clearly need an RFC.

Inequality in land and income ownership is negatively correlated with subsequent economic growth. A strong demand for redistribution will occur in societies where a large section of the population does not have access to the productive resources of the economy. Rational voters must internalize such issues. (Alesina, Alberto (1994). "Distributive Politics and Economic Growth"<sup>ⓘ</sup> (PDF). *Quarterly Journal of Economics*. **109** (2): 465–90. doi:10.2307/2118470<sup>ⓘ</sup>. Retrieved 17 October 2013. Unknown parameter |coauthors= ignored (|author= suggested) (help); Unknown parameter |month= ignored (help)) High unemployment rates have a significant negative effect when interacting with increases in inequality. Increasing inequality harms growth in countries with high levels of urbanization. High and persistent unemployment also has a negative effect on subsequent long-run economic growth. Unemployment may seriously harm growth because it is a waste of resources, because it generates redistributive pressures and distortions, because it depreciates existing human capital and deters its accumulation, because it drives people to poverty, because it results in liquidity constraints that limit labor mobility, and because it erodes individual self-esteem and promotes social dislocation, unrest and conflict. Policies to control unemployment and reduce its inequality-associated effects can strengthen long-run growth. (Castells-Quintana, David (2012). "Unemployment and long-run economic growth: The role of income inequality and urbanisation"<sup>ⓘ</sup> (PDF). *Investigaciones Regionales*. **12** (24): 153–173. Retrieved 17 October 2013. Unknown parameter |coauthors= ignored (|author= suggested) (help))

Should that be included in the Economic effects/Income inequality section? EllenCT <sup>(talk)</sup> 02:25, 2 January 2014 (UTC)

Survey

- Support** inclusion of the passage as a separate paragraph, to explain why income equality is a positive economic effect. EllenCT <sup>(talk)</sup> 02:25, 2 January 2014 (UTC)
- Omit** the paragraph in its entirety, as it does not even approach the subject of taxation, progressive or otherwise. Obviously off-topic and superfluous. Roccodrift <sup>(talk)</sup> 02:29, 2 January 2014 (UTC)

Closed →

## 2. Research Questions

- How can we build a tool that can make timely recommendations to an ongoing dispute to increase the chance of it to become resolved? The tool will spit out recommendations like..
  - Invite more expertised participants.
  - Change the discussion structure.
- Do people actually change minds during online deliberative discussions? If not, how can we make a better environment so people can have a chance to reach a better consensus through reasonable persuasion?

*At a higher level this project is about **deliberation and dispute resolution on the web**, not just about RfCs.*

### 3.Assumptions

- Here we assume that the unclosed RfCs remaining stale are a bad sign for the Wikipedia community since they end up with no resolution.

Unclosed RfC : bad sign for community



- Thus, we try to improve the chances of an RfC to become closed.
- The tool we envision to build will therefore predict the chances of an ongoing RfC to become closed and make recommendations based off it.

## 4. Jane's Main Contributions

- Created a novel dataset of **7,317** Request for Comments (RfC) from **4,997** Wikipedia pages.
- Analyzed the dataset to earn insights about the RfC process using various techniques including machine learning.
- Interview top RfC participants extracted from the dataset to understand tensions existing in the RfC process.
- Designed and implemented a voting feature in Wikum, a tool for summarizing online threads, in order to make it easier for editors to collaboratively close the RfCs.

# Quantitative + Qualitative study of RfCs

1. About the dataset
2. Which factors lead to an RfC closure?
3. Which factors affect time taken to close?
4. What are some problems existing in unclosed RfCs?



# 1. About the dataset

[Number of RfCs]

|  |       |
|--|-------|
| <b>Number of RfCs</b>                            | 7,317 |
| <b>Number of pages where RfCs were initiated</b> | 4,997 |

[Closed VS Unclosed]

| <b>Total number of RfCs</b> | <b>Number of closed RfCs</b> | <b>Number of unclosed RfCs</b> |
|-----------------------------|------------------------------|--------------------------------|
| 7,317                       | 4,083 (55.8%)                | 3,234 (44.2%)                  |

[Participants]

| <b>Number of total participants in RfC</b> | <b>Number of closers</b> | <b>Number of initiators</b> |
|--|--------------------------|-----------------------------|
| 15,213                                     | 760                      | 3,347                       |

## 2. Which factors lead to an RfC closure?(quantitative study)

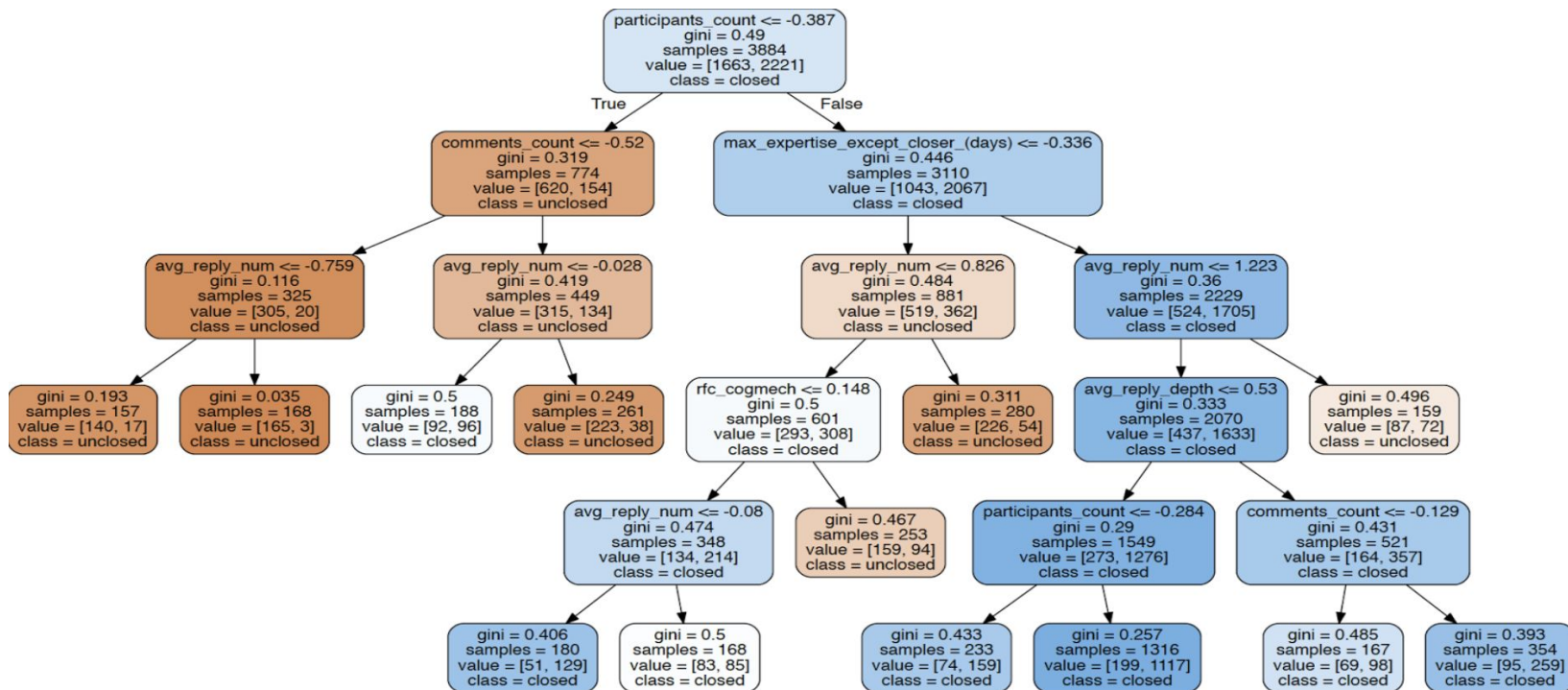
- **Baseline: 0.5668**, Train / test = 60 / 40, 10-fold cross validation, repeating 50 times, averaging the results
- Logistic Regression (LR), Adaptive Boosted Decision Trees (ADT), Random Forests (RF), and Support Vector Machines (SVM) with radial-basis function kernel (C=1, gamma=1/6)

|                                    |          |                             |          |                            |          |                                     |          |
|------------------------------------|----------|-----------------------------|----------|----------------------------|----------|-------------------------------------|----------|
| participants_count,                | 0.231429 | sum_participant_edit_count  | 0.02648  | rfc_incl                   | 0.016354 | two_weeks_recent_rev_count          | 0.007319 |
| max_expertise_except_closer_(days) | 0.142774 | rfc_positive                | 0.024583 | initiator_edit_count       | 0.015802 | one_week_recent_rev_count           | 0.006138 |
| avg_reply_num                      | 0.076779 | rfc_affect                  | 0.024422 | weighted_reciprocity       | 0.014729 | rfc_negative                        | 0.005761 |
| comments_count                     | 0.038736 | rfc_percept                 | 0.023326 | initiator_expertise_(days) | 0.013669 | rfc_swear                           | 0.004837 |
| avg_reply_depth                    | 0.03856  | avg_participant_edit_count  | 0.022529 | rfc_i                      | 0.012965 | two_months_recent_rev_count         | 0.004462 |
| rfc_cogmech                        | 0.036369 | revision_count_before_rfc   | 0.018646 | rfc_anger                  | 0.011539 | article_character_count             | 0.003794 |
| rfc_hostile                        | 0.031209 | avg_expertise_except_closer | 0.018632 | rfc_insight                | 0.010896 | article_word_count                  | 0.003369 |
| sum_expertise_except_closer        | 0.027798 | rfc_tentative               | 0.017407 | new_participant_ratio      | 0.009833 | one_month_recent_rev_count          | 0.002864 |
| sd_expertise_except_closer         | 0.026815 | rfc_excl                    | 0.01637  | rfc_certain                | 0.008531 | initiator_revision_before_rfc_count | 0.0024   |
|                                    |          |                             |          |                            |          | three_weeks_recent_rev_count        | 0.001875 |

*Features for predicting RfC closure with each importance after running tree-based feature selection.*

- **Attention RfC is receiving** (number of participant, comments), **maximum expertise of participant** (excluding closer if closed), **structure of discussion** (average reply number, average reply depth), **tones (cogmech, hostile)** are helpful for predicting RfC closure.

## 2. Which factors lead to an RfC closure? (quantitative study)



Closed RfCs tend to.....

- `participants_count` ↑, `max_expertise_among_participants` ↑
- `avg_reply_num` ↓, `avg_reply_depth` ↓

## 2. Which factors lead to an RfC closure? (quantitative study)

- Overall accuracy is about **77% (baseline: 55.8%)**
- RF has slightly highest accuracy
- Based on the features that are filtered out after tree-based feature selection.

|           | logreg   | adt      | rf       | svm_rbf  | dc_tree  |
|-----------|----------|----------|----------|----------|----------|
| accuracy  | 0.759846 | 0.784942 | 0.785035 | 0.776062 | 0.748919 |
| f1        | 0.796333 | 0.819332 | 0.822512 | 0.815756 | 0.788688 |
| precision | 0.766709 | 0.782043 | 0.773146 | 0.764286 | 0.754031 |
| recall    | 0.828338 | 0.860354 | 0.878801 | 0.874659 | 0.826689 |
| roc_auc   | 0.749285 | 0.773314 | 0.770577 | 0.760859 | 0.736928 |

### 3. Which factors affect **time taken to close**? (quantitative study)

- Why?: It is crucial for disputes to be timely resolved in any social computing systems.

| features   | $\rho$    | p         |
|--|-----------|-----------|
| Maximum expertise of participants (except closer)  | -0.105734 | p < .0001 |
| Number of total revisions made on the page where RfC is at   | -0.085853 | p < .0001 |
| Sum expertise of participants (except closer)  | -0.079067 | p < .0001 |
| Sum of participants' number of edits made during Wikipedia life                                    | -0.076892 | p < .0001 |
| Average expertise of participants (except closer)  | -0.075204 | p < .0001 |
| Number of revisions made on the page where RfC is at during the two months before RfC is initiated | -0.073167 | p < .0001 |

- Participants' expertise and activity on the talk page before the RfC is initiated seem to mostly influence the time taken to close.***

## 4. What are some problems existing in unclosed RfCs? (qualitative study → plan to find a way to quantitatively measure these problems)

### A.Reasons that are explicitly mentioned in the RfCs observed

#### 1.Initiator's fault

- Initiator was not clear with the wording of the request.
- Initiator's opening statement is biased (e.g. [canvassing](#)).
- RfC was inappropriate in the first place. Should have used another procedure.
- Initiator is a person who keeps bringing up issues (e.g. [forum shopping](#)). ⇒ Editors comment that the RfC is unnecessary.

#### 2.Participants (that are familiar with the topic, including the initiator)' fault: Bickering

Participants' bickering makes the overall **structure** and **content** hard for the new participants to read and comment.

#### 3.Other editors did not think there is a problem in the first place

No problem with the initiating statement and no bickering exists, but initial participants respond that what the initiator brings up is not much of a problem.

#### 4.Absolute consensus

A participant just writes down the visible consensus and takes the RfC tag off. ⇒ Some interviewees said this can be a good thing.

## 4. What are some problems existing in unclosed RfCs?

(qualitative study → plan to find a way to quantitatively measure these problems)

### **B. Other reasons that are just assumed from the shape of the discussion:**

- Discussion that has no comment or small number of comments
- Discussion that is long, with most of them not new inputs

#### **1. A topic that is unpopular**

**2. Not controversial enough** → Voting structure exists but pretty obvious what the consensus is.

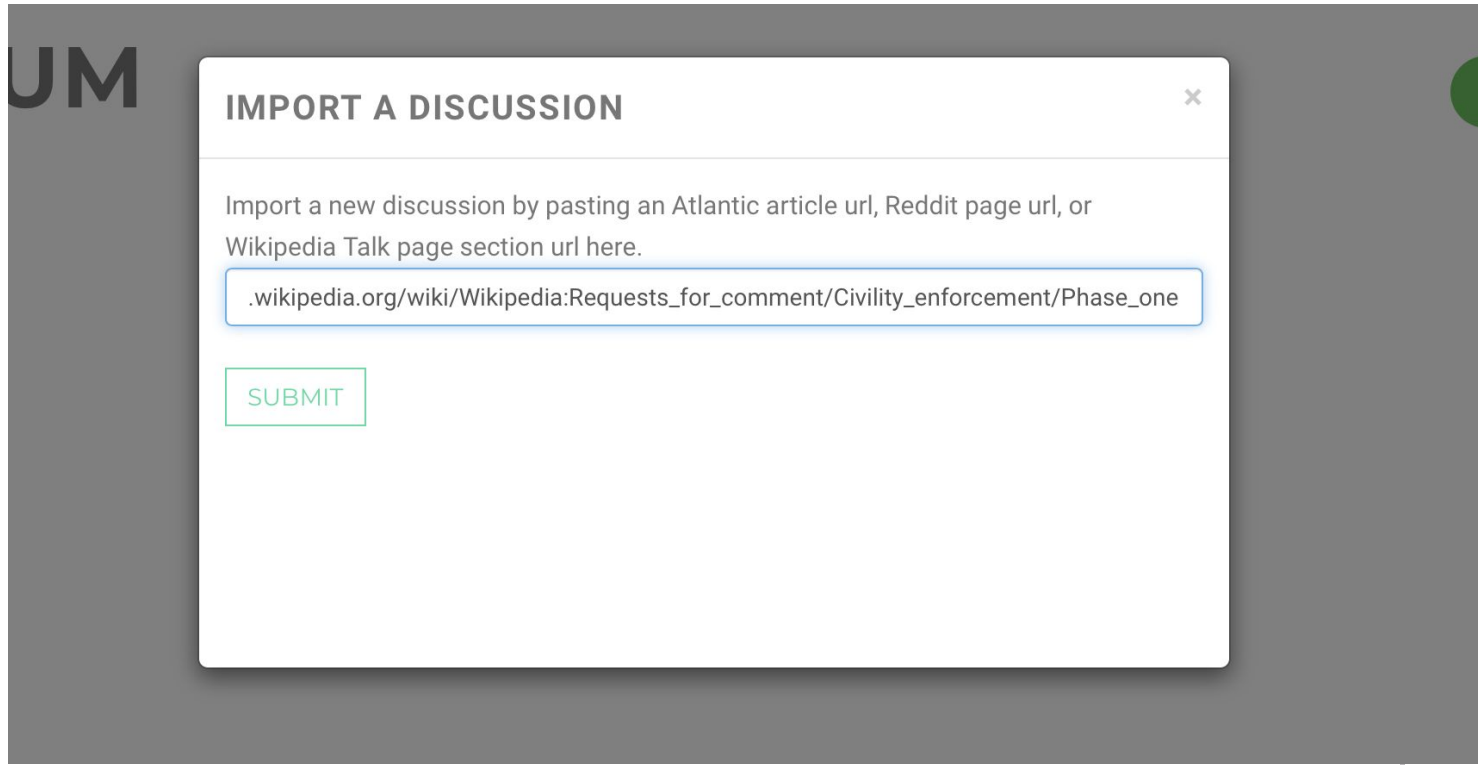
# **System-building: Developing Wikum**



# System-building aspect

- Contributions:
  - Fixed the codebase of [Wikum](#) (a tool for collaboratively summarizing online discussions made by Ph.D. student Amy X. Zhang) so it can import Request for Comments properly
  - Designed and implemented a **voting feature** into Wikum so editors can more effectively close contentious RfCs
- Acquired skills:
  - Django - web framework
  - Celery - distributed task queue
  - RabbitMQ - open source message broker

# System-building aspect



UM

## IMPORT A DISCUSSION

Import a new discussion by pasting an Atlantic article url, Reddit page url, or Wikipedia Talk page section url here.

SUBMIT

# System-building aspect

2% of this page summarized

WIKUM

Overall

Subtree

Summary

Edit History

Filter by tag

SORT ALL BY - ID

Highlight text

GET NEXT PAGE OF COMMENTS >>

ABOUT

CONTACT

LOGIN

SIGN UP

Talk:Planet Nine/Archive 4 - RfC: Images used for Planet Nine

EDIT ACCESS

## RFC: IMAGES USED FOR PLANET NINE

(Notifying previously involved editors: [Jehochman](#), [prokaryotes](#), [Serendipodous](#), [Fut.Perf.](#), [Ephraim33](#), [Nergaal](#), [Neutron](#), [Leitmotiv](#), [Kheider](#), [Wnt](#), [Nowa](#), [Itu](#), [Smkolins](#), [Tom Ruen](#) and [Jonathunder](#).)

I really think we need more input regarding the images used on this article. There have been a few previous discussions<sup>[1][2][3][4][5]</sup> but no clear consensus was demonstrated. I propose that we pool our collective opinions here and put things to a vote. Since there are 2 questionable images I'll split this into 2 subsections.

[Summarize Comment](#) [Tag Comment](#) [Mark Unimportant](#) [Show Hidden Comments](#)

COMMENT BY NAGUALDESIGN

### ARTIST'S IMPRESSION IN THE INFOBOX

Planet-Nine-in-Outer-Space-artistic-depiction.jpg

The infobox currently shows an artist's impression of Planet Nine (right), which appears to be closely based on [an image released by Caltech credited to R. Hurt \(IPAC\)](#). While artist's impressions may help to grab the reader's attention I do not think that it's becoming of an encyclopedia to reproduce them here and I propose that it be removed, or at the very least removed from the infobox. The only information it conveys are *basic assumptions*, which could easily mislead the reader. [A view of the Earth can be found here](#) in

# System-building aspect

## WIKUM

Overall

Subtree

Summary

Edit History

er by tag

hlight text

SORT ALL BY - # LIKES▼

Language ▼

Is Fermented Tea Making People Feel Enlightened Because of ... Alcohol?

EDIT ACCESS

### SUMMARY OF COMMENT BY DISQUS\_SJTSIXSRYG (2 LIKES)

There's good data on this already that the author did not reference.

"The outcome of this study included a study from my group which tested more than 100 samples, and found the majority contained more than just a trace of alcohol -- some products in excess of 2% alcohol by volume." [\[#41748\]](#)

1 0 | [View Original Comment](#) | [Edit Comment Summary](#) | [Delete Comment Summary](#)

[Tag Comment](#) [Mark Unimportant](#)

### COMMENT BY DISQUS\_2WZLUGNNEF (2 LIKES)

Funny thing is, Listerine mouth wash is about 12% alcohol.

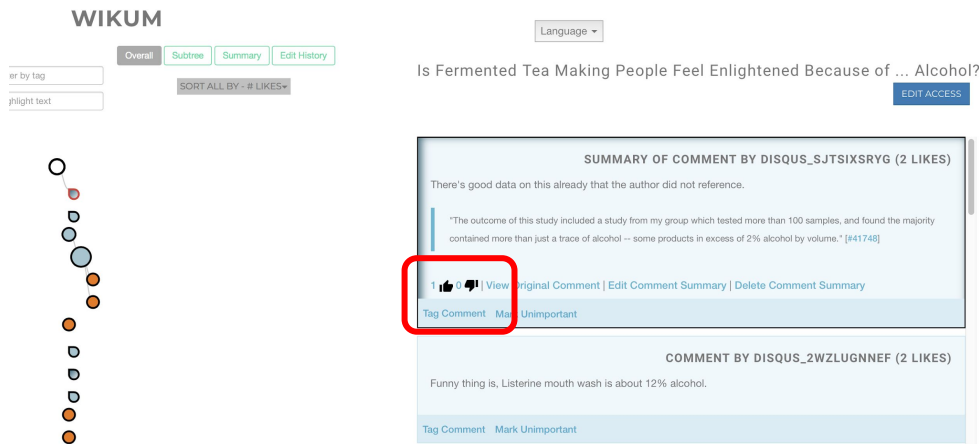
[Tag Comment](#) [Mark Unimportant](#)

# System-building aspect

- Interviews with top closers revealed that many closers **did not trust dividing up the work** during the closing process (e.g. writing summaries of comments) in collaborative closures.



- To resolve this I implemented a voting feature into *Wikum* so that closers can evaluate other closers' summaries of comments. I believe this can help the collaborative closings maintain a certain level of quality and relieve the distrust that many closers have.



The screenshot displays the Wikum interface. On the left, there's a sidebar with a search bar, a 'Highlight text' input, and a vertical list of colored circles. The main content area shows a page titled 'Is Fermented Tea Making People Feel Enlightened Because of ... Alcohol?'. Below the title, there's a 'SUMMARY OF COMMENT BY DISQUS\_SJTSIXSRYG (2 LIKES)'. The summary text reads: 'There's good data on this already that the author did not reference.' followed by a quote: '"The outcome of this study included a study from my group which tested more than 100 samples, and found the majority contained more than just a trace of alcohol -- some products in excess of 2% alcohol by volume." [841748]'. Below the quote, there's a voting section with a red box highlighting the '1' (like) button, a thumbs down icon, and a 'View' link. Other links include 'Original Comment', 'Edit Comment Summary', and 'Delete Comment Summary'. At the bottom, there's a 'COMMENT BY DISQUS\_2WZLUGNNEF (2 LIKES)' with the text 'Funny thing is, Listerine mouth wash is about 12% alcohol.' and links for 'Tag Comment' and 'Mark Unimportant'.