# Normalisation and Improvement of the Soyboorus Media Category System

By Yaroslav Yamamobich

# Terminology

**Booru** - A media hosting site that is primarily focused on hosting images, but also sometimes videos, flash files and the rest depending on implementation, typically running danbooru software or its forks.

**User** - a booru user that has no specific privileges and can contribute to the site with uploads and tag categorisations and search its contents using the tag system provided by the booru.

**Convention** - In this context, guidelines that enforce the booru styling, implemented either in form of program filter or moderation, usually created for the increase of tag usefulness.

**Cascade** - In this context, to implement all the ancestral tags down to the parent zero, for example "subvariant:pyramid > subvariant:circle > variant:feraljak".

**Yamamobich system** - System proposed in this paper

**Complex tag** - A tag that gets interpreted into several tags defined in the complex tag and itself.

**Alias tag** - Complex tag that doesnt resolve itself

**Namespace** - Tag group in the Yamamobich system similat to boorus categories.

**(Media) Subject** - Character or thing in the media that can be described with features that are specific to it.

# Abstract

Soybooru has a major malfunction regarding its categorisation system inherited from the predecessing booru software which was notintended for the Soyteen demands. Subvariant categorisation forces people to cascade categories by hand and reference its ancestor, and reference it regardles of if features of it are present at all.

It is then later argued that the soyjak categorisation is not a variant hierarchy, but rather a concatenative type system, and that the strict hierarchical subvariation should instead be reformed into a system where subvariation is rather a universal attribute, than a type specific to its parent variant. Paper also proposes an elimination of the writing of the common attributes altogether, by proposing inheritance.

The system would then be reformed into display tags, that are verbose and are stored for each picture. Display tags would be shown on the individual posts and are going to be used for identification within a database. In addition to the "traditional" tag set there would be then abstracted tag set that an uploader would submit, which is going to have normalised set of tags that prevents categorical redundancy, where abstract classes are assigned their common features, the poster would then have the option to

exclude those predetermined tags with subtractive set operation, instead of having "no-" tags in the tag set.

The paper will later describe how to implement the said system in the existing vibegqded informational structure. It would propose gradual or partial change, where select features could be added, starting from the ones that can be already partially implemented without any coding, like arbitrary namespaces, in which you would just free people from strictly defining groups and concatenative categorisation, where you untie subvariants from the variants in the convention.

# Relevance

Within the Soybooru theres a section in which a user can view media, and most importantly the media displayed has a "tagging" system that puts it into categories by which the user can then filter the displayed media based on the search term that is composed of the tags in question and set operators. Tags are also included in their own subcategories.

Though it is a major case of SNCA, but the similar tagging system exists in the rest of the booru imageboard sites, and it is important to note that the system have existed for longer than most of the Soybooru users mom hoes and was never improved upon to meet the specific demands of Soyjak booru. While the tagging system does a mediocore job at searching junko foot job pictures but the actual creative community that is Shitteens generate content that goes beyond intended boorus scope of categorisation with some of its posts being pure fucking categorical nonsense[1].

# Issues with the current system

As was said in the previous section the current tag system shows alot of problems of categorical redundancy in the context of the soybooru.

The current system is used for categorising soyjak creations that begin their proliferation and modification cycle from a variant which is either a direct/indirect trace from a source picture or an original artwork (noone does this). Those variants later recieve minor edits, usually the ones that are translatable to several variants after which it is necessary to create a tag for finding the variants in question.

After that, based on the existing variants, users would create edits of them, by modifying those in a maner that is specific to a variant, making it unlike the rest of edits not translatable to the rest of the variants, which would make it in category relation similar to variants and are therefore labeled as Subvariants. Subvariants can then recieve their own edits and their own subvariants in the booru with no limit. Becuase a

1 https://soybooru.com/post/view/242590

subvariant is made from a variant or a parent subvariant, an uploader moderator would then need to cascade its subvariants when defining the tags for the media.

The separation between the subvariants and edits are made arbitrarily for the convenience and are not based on real limits, which might then result in categorical problems.

## Variant Theseus Ship Problem

Theseus ship is a category problem in which there is an original ship that is gradually modified of its parts, which would then raise the question of if the ship is still genuine or not, based on the fact that most of the parts, that define the ship, are not present.

The problem then translates to the variant -> subvariant system, where the subvariants would gradually change its parts, eventually reaching a point where most, or sometimes none of the original parts remain in the subvariant, yet the system necessitates to define every predecessor.

First, funny and more intuitive but not general counterargument against the "Variant > Subvariant" system is a hypothetical subvariation from feraljak to cobson. Lets suppose that there is a subvariant, in which the variant:feraljak is modified by changing his features to cobson. Later, this subvariant could recieve modification which would replace even more parts of the feraljak with cobson parts. At the last step, we are going to completely replace feraljak parts with cobson parts in such a way that we reach the original cobson image one to one, which would imply that cobson is feraljak subvariant which it obviously fucking isnt. Therefore, current category system is prone to contradiction.

Second, more general counterargument starts with defining what the initial demand for the category at hand is.

A user, when searching for a subvariant, would demand all the verbose cascading variants because they can possibly not know the subvariant itself, but assume its ancestor based primarily on the visual appearance, as it is the primary identificator for the variant/subvariant after its assigned tag on booru, The uploader also primarily tags the media based on its visual appearance. It is an intuitive way of determining the origin of post, rather than learning the "convention". If the systems purpose is to provide user with a set of tags to make it easy for him to fetch/categorise post it is therefore primary for the tags to represent the visial queues, and not arbitrary history of the subvariants development. Therefore, having the origin variant when its visually indeterminable is useless for a tag system that focuses on the visual queues.

## Subvariant hierarchy problem

Based on the previous argumentation, on the fact that the tagging system should be built upon visual features, therefore there can be argued that the necessary variant

hierarchy is therefore redundant, as it is also not inherently representative of the visual queues by which the user would try to find the media.

The property of subvariant, which separates it from the edit, is also completely arbitrary. Additionally, a user might do, though individual and not "scalable" but still similar subvariation for each jak which would create repetitive subvariants, which shouldve been otherwise categorised like the rest of edits that are done across variants, as just edits to be appended to the variant, instead of strict hierarchy.

### Inferable tags problem

When categorizing media, a poster would usually have to describe every single feature of the post. There is a crutch solution to this which involves copy and pasting of the tags onto the media from its origin, which raises the question of if defining the individual tags for those types of posts is even necessary.

When defining a post on the booru, a user should be able to describe the picture in the most intuitive way to how it is differentiatable from the rest of the pictures, so the particulars that are inherited from the identifiable tags are not necessary and are only adding to the unnecessary categorisation work.

In the Soybooru, there is medias in which theres several variants that are then counted up in the "soyjak counter" going 1, 2, 3, 4 for particular amount of soyjaks to multiple which is used to describe posts in which there are 5 variants or more. This is also redundant as it is inferable from the amount of individual variant groups that can be described by the uploader.

### Subjective category problem

Booru obviously cannot provide concrete categorisation for properties of media that involve subjective opinion, like if the post is either ias or not. And even if the "convention" is created on the mater of categories that were previously believed to be subjective, it runs into a similar problem of convention being unintuitive for defining or searching media on the booru.

This is not entirely fault of the system, but at the same time it does not provide any particular tools of determining such categories either.

## Proposed system

Proposed system aims to provide a category system that fixes the present category systems problems and aims at providing tagging system that means the demands of users of the booru and improve process of searching and categorising the media. The system will be built from grounds up, though it is assumed that the previous category principles are already known.

To differentiate between the previous and proposed model, the first will be referdd to as "Booru categry system" and second will be refered to as "Yamamobich category system" because i need to drag the shitty joke.

The system is also syntactically going to mostly resemble and be compatible with the current system as the users, again, cannot be bothered to "stick to convention" for their convenience.

## Syntax

Tags are character strings separated with spaces.

Uppercase characters are ignored, and are striped for the display tag.

For the sake of convenience, the underline character can be used as a space character for the name of the tag.

Characters that are not letters OR underline are going to act as special characters which would either perform actions or form special categories. They can either retain within the provided tags in the stored string of tags or be interpreted by the program.

## Display and submit tags

For the display and storing of the tags for search in the database and for the submition of tags in the search or edit form there will be separate strings of tags.

Display tags are final tags, they are the compiled version of the submit tags that is otherwise more abstract and contains set functions.

## Arbitrary namespaces

In the model, a tag can be assigned arbitrary namespaces at an arbitrary amount. This creates the additional function of differentiating, for example the variant arrays by putting the features specific to a variant to its own numerated category, which would look like this:

```
1:variant:cobson 1:red_skin 2:3:variant:feraljak 2:blue_skin
3:subvariant:cryboy 3:green_skin handshake
series:incomprehensible_screeching
```

The namespaces resemble the categories of the booru category system, but are instead free to define and are concatenable without particular order, with the "origin" tag being the rightmost separated string. Such a system makes it determinable which subjects in the media have which attributes, which would otherwise be undeterminable if the specifying namespaces werent present.

Also any tag could be assumed to have "their own namespace" (Cobson has a `face:cobson` which is cobsons face, duh), this is important for later operations

## Unique namespaces

Since the Yamamobich system allows the differentiation of the tags, there can be implemented a feature that would otherwise not be posssible, which is namespace tag restriction to be unique to a subject namespace, which would prevent a submitter from assinging both black and white color to a medias subject for example. This on the other hand, unlike arbitrary namespaces should be manually defined when necessary (as the "uniqueness" of a property is subjective to the situation, for example, thouhg it is immediately obvious that soemthing like skin color should be a unique namespace, but, there can be the case where the the subject of the media would have both colors and it would not be possible to tag both due to limitation)

## Post tag inheritance

In the submission tag, instead of switching tabs and copy pasting tags a select post can be refered in the tag and manipulated upon with a special namespace denoted by hashtag character prefix

*Though it is out of the scope of this paper there an also be proposed a feature where a post can be assigned an identifier so instead of putting in post id you can refer to it by a keyword.*

## Set definitions

In the Yamamobich system, there can be defined "complex" tags, which would contain itself and appended tags, and then unpacked into the display tags by the program. This is the primary feature to resolve Inferable tag problem.

To compensate for the excluding cases where the media described by a complex tag doesnt have its appended feature, that feature can be excluded by an excluding submit tag, so the **`variant:feraljak -glasses`** would define a feraljak with all his features except glasses.

Also when defining the tags, complex tags can be joined (besides regular AND operation) where complex tag can be subtracted of other complex tags categories (**`t-t`**) or replace the namespace of the complex tag with other tag ones (**`t.n-t`**) (The operations obviosly are not associative!).

**`variant:cobson.face-subvariant:cobson_front laugh ai_generated meta:video`**

This allows for definition of otherwise very redundant and hyperspecific selectors in a simpler maner.

Important lead of the set definition is the redundancy of the subvariant cascade with the system (and therefore solution of the Variant theseus ship problem and Subvariant hierarchy problem), as you can just define all the parents with it, it also removes the burden of needing to know what the actuall fucking origin variants of something like vlodson is.

### Convention tags

To resolve the subjective category issue, without going into the barbaric measures of fun policing, there can be implemented convention tags that can be implemented by conventions, which are groups that would either operate either on blacklist or whitelist principle.

Whitelist principle would allow only people within a group to issue a specific tag, which can be used on something that is thought to be very abusable and which would necessitate gatekeeping, it can for example be used by group of people that archive content or (You) to be differentiatable from other fartists.

Blacklist principle are otherwise conventions that exclude particular people instead (which makes any user able to use the tag by default) and can be used for soemthing given to people assuming good faith in them in contribution, which would fit for something like IAS category.

The convention tags are going to be, in the Yamamobich system, denoted with at character prefix.

# Implementation

Important factor, which will hinder the implementation of the papers category system for the booru, is the considerable lack of coders in soysphere bruh. Therefore the implementation is better pulled off partially, maybe even in steps to achieve the described system somewhere in the future.

First suggested step is to implement the changes that can be done on convention, as it is as easy as changing guidelines, which is to make categories more arbitrary, and focused on the visual fact rather than based on a history of the variants, which is shit nophono cares fan about.

The described system can be reformed into simple ones for easier implementation, for example "convention tags" can just be made into "collections" with set managers, which is though not autism affirming, solves the same problem at hand.

To finish this document, is is also asked to atleast consider the critique described in the paper, even if none of the proposed changes in any form would be implemented.

you, yes you, if youre going to say that this is ai generated i will fucking find you and kill you and rape you and fuck you also yo mom hoe fucked in the sleep.