Tyler Procko – *Semantic Science*
86130 Williams Avenue
Yulee, FL 32097
March 14th, 2024

# Semantic Science

- Ontology Engineering & Research Publication -

www.semanticscience.us

## Plan

The present document is a formal plan for the rest of the project. I need to express my thoughts as clearly as possible. As the taxonomy has evolved to version 0.2.2-alpha through several iterations, it has reached a point where its usefulness is apparent, and therefore it should be integrated with the Hooray sites. The current interest in using the web post taxonomy is for two primary purposes, with an added third:

1. Adhering all new (future) posts to the taxonomy: this amounts to a "teaching" or "training" method for all Hooray authors, effectively forcing them to adhere to the taxonomy as closely as possible; the method for doing this is outlined later
2. Converting all old (past) posts to the taxonomy: this is a combination manual/automatic effort for converting the categories of all past posts to the new taxonomy, which is likely of less importance and can be done in due time, as opposed to adhering new posts to the new taxonomy, which all authors should made aware of as soon as possible
3. Hosting the taxonomy publicly, in the style of the W3C: not the pressing concern, but of interest for the credibility of Hooray and for demonstrating transparency in data

## WordPress

Initially, it was thought that a script-driven approach through the WordPress API was the best way to upload the taxonomy, convert old categories, etc. However, as investigation revealed, there are a number of relevant WordPress plugins that can perform what is needed, i.e., everything can be done within the WordPress UI, which saves time and effort.

The first step is to import the taxonomy itself. A new taxonomy can be created, e.g., with the Taxopress plugin, or the default WordPress category taxonomy can be used. The latter option is easier, but may present issues if category terms already exist on a live site, e.g., *Christmas*. Ideally, the terms can be imported into a new taxonomy, all posts adhered to it, and then the new taxonomy migrated to be the default Wordpress taxonomy. I need John's input here.

For now, to import the taxonomy, the easiest way is to use the Category Import Reloaded plugin. I have a script to output the entire taxonomy from the OWL file into lines separated by the / character. Paste the output text content into the plugin and run it. Now, navigating to the Posts > Categories tab, all category terms are visible. If desired, all items can be selected, then Bulk Actions changed to Delete to remove them for testing. Anyhow, if it is enabled, posts can be categorized normally by authors.

The WP All Import plugin could also be used, but their interface is more complicated for just importing the taxonomy and I do not think it is necessary in this step. Another option is WPSheetEditor. See the later

section titled *Keeping the Taxonomy Updated* about this. For now, let's just assume the taxonomy is set and ready to go.

**Adhering New Posts**

The most forward-looking problem is getting Hooray authors familiar with, and able to use, the new category taxonomy to classify their posts. We have to establish a repeatable process they can follow every time they post; i.e., they must all be "on the same page". The current issue is that they all use their own terms to categorize – this proposed process would remove that issue. Following are the proposed rules for categorization after authors have completed a post for publication:

1. Authors must select at least one, but no greater than two[1], region terms, e.g., *Orlando*
2. Authors must select at least one, but no greater than three, other category terms, e.g., *Holidays*
3. Authors should select the most specific term possible, when applicable; for instance, if a post is discussing Christmas tree decorations, the author should select *ThingsToDo/ReligiousHolidays/Christmas/ChristmasTree* along with *Christmas* itself. It is important to not only utilize the taxonomy to its lowest (or deepest) terms, but to consider SEO, as *Christmas* is likely an important category and search term. A third category could be *Seasons/Winter*, or *Home/Crafts*, at the discretion of the writer.
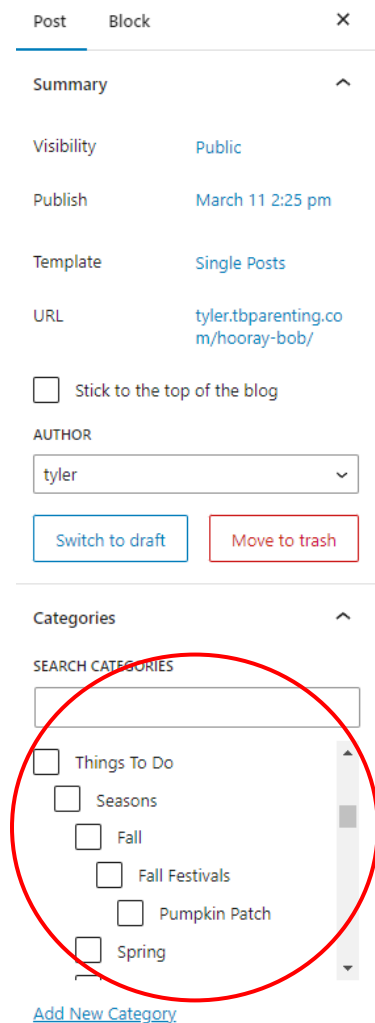
This allows for a total of 5 categories (2 possible region categories + 3 possible other categories) per post, which is not too many or too little; but the exact number can be debated. Allowing more regions or other categories would make post categorization too bloated. More is not better, less is not better; precise is best. Finding that precise value is important. In any case, it is John's choice, in the end.

There are a few ways for the authors to pick which categories their new post should be placed under, after carefully considering the content of the post and potential SEO concerns:

1. Visually, using the taxonomy image, by zooming and navigating until the categories are selected
2. Within WordPress itself, by scrolling through the taxonomy hierarchy or searching for the term desired
3. Using the GPT post categorizer as a recommender (remember this?)
4. Some WordPress plugin that automatically categorizes against your taxonomy (this probably exists with GPT being so popular)

This is preliminary. Ideally, this process would be formalized in a training video for the authors to be taught with, or perhaps even live demonstrations with John and/or I.

---

[1] This allows for posts which discuss more than one location to categorize under those locations. If more than two are needed, the "region areas" can be used, e.g., if an author wants to categorize with *Ocala*, *Orlando* and *Sanford*, she may just want to categorize with the higher term, *CentralFlorida*.

It may be possible to simply import the new taxonomy as a sibling (or shadow?) taxonomy, have the authors use that (I am not sure how to allow this in the admin UI), and then later roll the sibling taxonomy into the main WordPress post taxonomy.

## Re-categorizing Old Posts

In total honesty, I do not see the conversion of old post categories to the new taxonomy being possible in a perfectly automated fashion. WordPress's UI and API are both pretty terrible, so some manual effort is probably needed. What I am trying to avoid above all else is leaving any posts uncategorized, even for a moment, so we do not injure SEO. Below is the proposed process for converting all old post categories to the new taxonomy.

1. Using WP All Export, grab all posts from a given Hooray site (this avoids the issues of webscraping as we have discussed previously)
2. Run my GPT post categorizer over every post and output 5 recommended categories (or more if desired) into a CSV spreadsheet
3. Either:
   a. Give this spreadsheet to a few Hooray writers and have them go through every post and update the post categories as they deem fit, using their judgement and knowledge of the new taxonomy
   b. (Very theoretical and not even sure it is a good idea because GPT can make mistakes) Merge the content of this spreadsheet with the WP All Export CSV containing all posts, replacing the values in the categories column with the new ones, then upload the entire site at once, effectively "replacing" the old content in one go

In either case with the third step, recategorizing is not difficult. The decision just has to be made.

## Keeping the Taxonomy Updated

Keeping the source ontology and the WordPress taxonomy equivalent in version should ideally be done automatically. The Category Import Reloaded plugin will not import terms that already exist, so if the source ontology just has new terms added, this plugin can be used without issue. However, if the order of terms is changed, or terms are re-named, automatically updating the taxonomy in WordPress is not directly possible, and will require manual fixing. This is why it is absolutely essential to ensure that the ontology is as good as possible over more iterations with Hooray authors, so we are not making massive changes down the road. See the *Iterating with Laura and Others* section of this document.

Anyhow, a plugin that might work is WPSheetEditor. It requires a specific CSV format for the taxonomy, but I can write a script to do this. Once in this sheet form, categories can be edited individually or in bulk, which may make keeping the taxonomy up to date with the source ontology easier.

For simple edits to the taxonomy order, the plugin Category Order and Taxonomy Terms Order (once installed, it's found in Posts > Taxonomy Order) is great for drag-and-dropping order rearrangements of the taxonomy.

## Open Ontology Hosting

The basic process for hosting our Web Post Ontology (WPO for short 😊) is outlined below:

1. Generate static HTML page for it (I do this instantly with PyLODE)
2. Pick an exact ontology URL, e.g., *https://hooray.media/ontology/wpo/version/0.2.2-alpha*

3. Make a page for it in your site (if WordPress, you know what to do)
    a. Ensure this URL matches your actual base ontology URL
4. Upload the HTML to that page
5. Enjoy

I use the # character to separate terms from the ontology URL, e.g., …#Christmas, …#Birthdays, etc., so when someone actually goes to *https://hooray.media/ontology/wpo/version/0.2.2-alpha#Christmas*, it opens the ontology file right to that line. This [old guide](#) from the early days of the Semantic Web outlines it pretty well.

### Tags

As for tags, we have discussed them as being "instance" data, e.g., actual places, people, etc., as opposed to categories, which are general terms that *may* be instantiated. WordPress does not allow its built-in tags to have a hierarchy, and I think that is fine.

In a previous stint, I did scrape several hundred Hooray posts to extract tags, but the method was rather primitive. A real method for tag extraction would invariably involve either regex work to locate "named" things, e.g., those starting with capital letters, or, realistically, parsing with GPT to extract tags as we have defined them to it. I can work on this much easier, and without webscraping, with the WP All Export plugin now, if desired; but the taxonomy work is my primary concern.

### Iterating with Laura and Others

The final step is to rigorously ensure that the taxonomy is complete enough to use. There are currently 272 classes in v0.2.2-alpha of the taxonomy. Given the taxonomy visualization, Dani Meyering had immaculate feedback that was integrated immediately, particularly in the Holidays sub-tree. It is likely that Laura Byrne, and other Hooray authors, could provide their own insight, and the taxonomy could become more complete and correct. So, more feedback is needed, and then the taxonomy can proceed to its first major release, v.1.0.0-beta. If it is evaluated to be useful on a Hooray site (ideally the smallest with the fewest posts), it can be released as v1.0.0 and integrated with all Hooray sites.

In particular, I believe that the *Family* and *Home* sub-trees could use serious improvement, as they are too shallow. These two need attention, as well as the *News* and *People* sub-trees, which I think are less important.

### Conclusion

The plan is pretty straightforward. For new posts, we simply train Hooray authors to follow a set of rules in picking the new terms. For old posts, they are re-categorized semi-automatically (or fully automatically, if we trust GPT). In any case, I have a few concerns that maybe you can address:

1. Should the new taxonomy take the place of the default WordPress taxonomy, or should it be a sibling/shadow taxonomy? Does it matter for SEO? Does it matter for your site functionality? Does it matter for the ease of your authors' use? Can a sibling taxonomy be shifted to be the default taxonomy later, if desired?
2. Another concern is the "mapping" of taxonomy terms. Many of the terms currently used by Hooray authors, e.g., *Christmas*, are present in our new taxonomy, but many are not. WP All Import seems to have some functionality for [mapping](#), but I really cannot grasp their interface, and it is also not a bulk mapping, so I do not think this is viable. I suppose the question is: do we care? If we intend

to move all authors to the new taxonomy, and then re-categorize all old posts later, is it even important to consider mapping the old categories?

3. Let's get more feedback from the other authors, especially Laura, because we need the taxonomy to be as fleshed-out as possible before releasing it as v1.0.0-beta; can we send out a blanket email with the taxonomy image and some instructions for what we are looking for?

4. Let's get a license for WPSheetEditor, it looks useful for keeping the taxonomy updated over time

I apologize for the delay. I needed to step away to think about the approach for this. If you are in agreement, let's advance with the new post taxonomy adherence first, by iterating a few more times to v1.0.0-beta of the taxonomy, importing it to one or more of your sites, and training your authors with a video or live demonstration. If we deem this to be viable and useful for the authors, we can execute the plan for re-categorizing all the old posts. Please let me know your thoughts!

At your service,

Tyler

# Semantic Science

- Ontology Engineering & Research Publication -

www.semanticscience.us