

Tyler Procko – *Semantic Science*
86130 Williams Avenue
Yulee, FL 32097
December 1st, 2023



- Ontology Engineering & Research Publication -

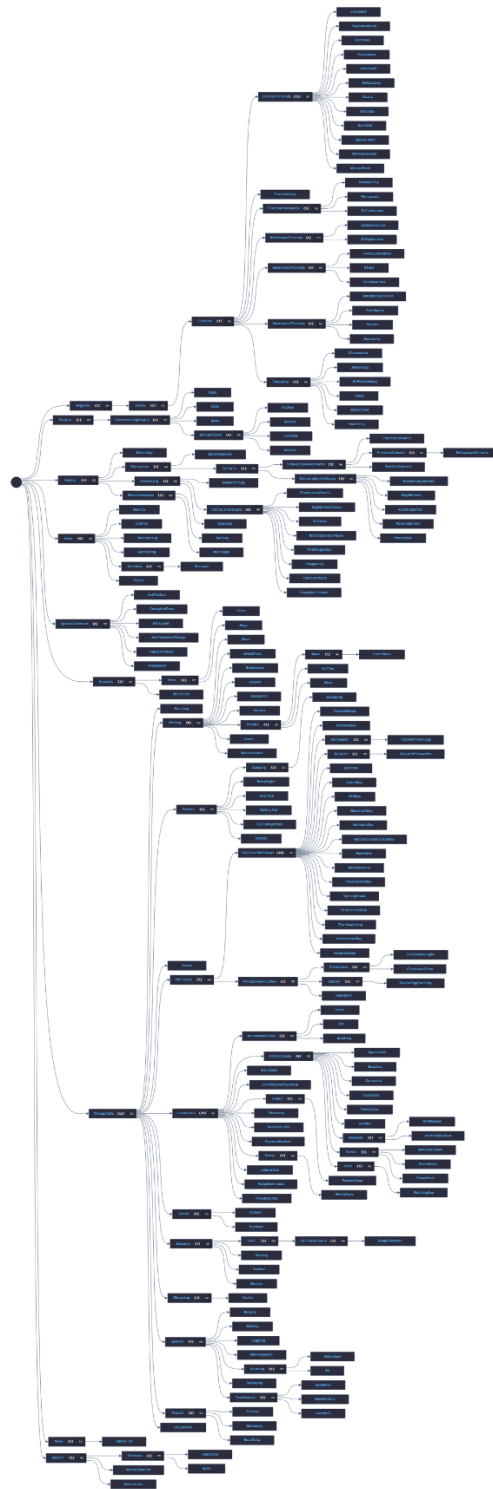
www.semanticscience.us

Updated Taxonomy

Some interesting progress has been made on the taxonomy, per our last meeting. The exact changes are detailed below:

- Added top-level class Family
- Moved Parenting under Family
- Moved Education under Family
- Added Decorating under Home
- Added Schools tree under Education
- Added Relationships tree under Family
- Added Occupations tree under Interesting People
- Added Daycare under Parenting
- Moved Birthday under Family
- Added Churches class
- Broke apart the Travel class because it is semantically incorrect
- Moved Road Trip under Travel
- Added Dating under Relationships
- Added Nutrition under Health
- Added Gyms under Fitness
- Added National Park under Park
- Added Daycare under Parenting
- Several renames from singular words to plural
- Added Gardening under Home
- Added Garden under Locations/Attractions
- Added Team Sports tree under Sports
- Added top-level class Work
- Added top-level class Special Content and several sub-classes
- Added COVID-19 under News
- Significant changes and additions to the Regions/State/Florida tree
- Added Exercise under Fitness
- Added Jogging under Sports
- Added Wildlife under Animals
- Added Motorsports under Sports

The size of the taxonomy has increased approximately 30%. This is reflected in the JSONcrack visualization:



I can no longer give you a link to the interactive version because the taxonomy is too large to view without a JSONcrack account. In any case, please see the .png file bundled with this progress report and zoom in/move around as you please. I also included a full-res .svg image that should open with Microsoft Edge.

Automatic Categorization

As for the automatic categorization of webposts, I am still engineering an API-based solution in Python that you can potentially run in the back-end, but for right now, if you have a ChatGPT+ account (\$20 bucks a month, you should get one!), you can access a GPT I trained on the taxonomy files.

- [John Byrne Classifier](#)

Sign in, copy/paste the content of one of your webpage posts, and get suggestions for categories based on our taxonomy! Still very preliminary, but very promising!

The screenshot displays the John Byrne Classifier web application. On the left, a browser window shows a restaurant review titled "Brunch Buzz: Carmel Kitchen & Wine Bar" by Samantha Rosenthal, dated March 8, 2017, and updated on August 27, 2018. The review is categorized under "EATS" and features a photo of a dish. On the right, the classifier's output is shown, listing applicable categories from a taxonomy: Restaurants, Brunch, Dining, Orlando, Drinks, Food Items, and Recipes. Each category is accompanied by a brief explanation of why it applies to the post. The classifier also notes that the categorization aligns with the taxonomy's focus on parenting and dating lifestyle content, specifically targeting dining and brunch experiences in Orlando.

As of right now, this can essentially be thought of as a constrained recommendation machine for your content writers. With access to Wordpress' API (and when I can get OpenAI's beta features to work), this could potentially be totally automatic. You can set up business/organization accounts with ChatGPT, so maybe this is something you can just share with your content writers, if that would be easier. Still, it needs some more fine-tuning (the exact behavior of a trained GPT is stochastic, i.e., it takes some revising to get it working perfectly), and some more work on the taxonomy, before it can really perform perfectly.

Tagging

Outside of these developments, the only remaining challenge is tagging, which, as we roughly agreed upon, are distinct from categories, in that categories are not "tangible" classes of things, and do not change much, such as the category *Restaurant*; whereas, tags are likely real "instance" things, and may change a lot, such as restaurant names, such as *Louie's Bar on 5th* or some real person's name.

I started this process by (as you've seen) annotating your sitemaps Excel file. After this iteration of the taxonomy, all the terms that really remain in the annotated Excel sheet are too specific to be categories, so

they must be tags. But I did not think this was enough. So, I wrote a script to grab all of the links from all of your post sitemaps (all 10 of them); I then took each group of links and put them into GPT (the default model, not yours), asking it to give me as many tags in plaintext as possible, to supplement the (often) hyper-specific tags some of your writers employ (mostly on the ODNG site). Here's a sample:

- Playdate Orlando
- Cheap Phantom Tickets
- Key West Flights
- Spa Date
- Valentine's Day Planning
- Shamu Valentine's Day
- Valentine's Cooking Class
- Valentine's Happenings
- Chillounge Night
- Amazing Race Date
- Festival of Chocolate
- Vine and Dine Series
- Trivia Night
- Hamburger Mary's
- April Happenings
- Loews Events
- Orlando Science Center
- Cuban Pig Roast
- Prix Fixe Menus
- Cheap Eats Pointe Orlando
- Vegan Date Night
- Orlando Limousine Scavenger Hunt
- Sushi Making Class
- Roy's Cooking Class
- Culinary Canvas
- Enzian's 007 New Year's Bash
- Michael Andrew's Holiday Spectacular
- New Year's Eve Ideas
- Starry Nights
- Cocktails & Cosmos
- John Gottman at Rollins College
- City Scramble
- Valentine's Painting Class
- Culture and Cocktails
- Monthly Wine Lounge
- Cheese and Wine Class
- Winter Park Sip Stroll
- Music and Mojitos
- 18+ Bookstore Date Night
- Orlando Arts Sample
- Date Night Idea Deck
- First Fridays Art Stroll
- Battle of the Parks
- Festival of the Masters
- Gaylord Palms Ice
- Holiday Ice Skating
- Popcorn Flicks
- Disc Golf
- Holiday Hoopla
- Paint & Sip Experience
- New Year's Eve Part Two
- Winter Park Sip Stroll Thursday
- Wine Dinners
- ArtsFest
- Aphrodisiacs Myth or Reality
- Nude Nite
- Love Your Shorts Film Fest
- Orlando Date Night Awards
- Fine Print
- Grand Prize Revealed
- Moon Zoom After Dark Adventure
- Good Dining.com
- Chefs Table Experience
- The Downtown Pour
- Wine Bar Dates
- April Date Night Calendar
- Pasta Making Class
- Dining in the Dark
- Orlando Restaurant Week
- Orlando Fringe
- RanGetsu
- 100 Date Night Ideas
- Monthly Wine and Cheese Tastings
- Advertise
- Retro Game Night
- There Will Be Words
- Jana Banana's Date Night Picks
- Orlando Food Tour
- Trivia Date Nights
- White Peach Sangria
- Cookbook Club
- WFTV's Daralene Jones
- Dark Knight Marathon
- Dinner at Primo
- Summer Nights Orlando Science Center
- Ritz Carlton Foodie Weekend
- Orlando City Soccer Date Night
- Bohemian After Dark
- Friday Night Food Truck Event
- Wine Spirit Walk
- Farris and Foster's Famous Chocolate Factory
- Date Night at the Picture Show
- September's Date Night Calendar
- Bash for Books
- Orlando Arts Samples
- Progressive Dinner Date Night
- Meditation and Libations
- Saturday Night Dueling Pianos
- Global Peace Film Festival
- Orlando Magazine's After Dark Ideas
- Date Night Ceramics Class

Obviously, there are many strange tags, but still some applicable things that we can (1) possibly elevate into categories and (2) dump into bins of tags for your writers to choose from instead of making new ones. I generated thousands upon thousands of them. Lots of work to do there.

Future

First, I am going to put this all together into an organized GitHub repository so I can just point you to the most updated work without relying on email, from now on. GitHub also has a built-in project board that is immensely useful for tracking project ideas, plans and issues, as things progress. It is not difficult to use, and can be set private, so only you and I can access it, until you choose to release things, if you ever do.

Second, I intend to mop up as many tags from the extant sitemap Excel sheet I've been annotating and the GPT-queried tag lists (which is ongoing). I will sift through these to find relevant categories that might have been missed, and then the remaining terms, too specific to be categories, I will start to prune and bin under certain categories. You mentioned that Wordpress allows many content taxonomies; this is great. As I envision it, you would have two large taxonomies: your categories, which you would ideally display across all your sites (whatever set of terms you deem most important, of course); and your tags, which is a bit more "wild west", but still controlled under your category taxonomy. While both are proprietary, I think your tags taxonomy shouldn't be released to the public, and the category taxonomy's status is up to you.

Other things of interest and/or general notes:

- Translation of this taxonomy into a functional Wordpress taxonomy... I'm not sure what format, or through what interface, Wordpress expects content taxonomies to be made (I mean outside of their web interface); there is likely an API call that can be made, and if so, it's just a matter of a Python script and access to your organization to do so
- I noticed that some of your post URLs are prefaced by year/month, e.g., 2023/12; only *some* posts are, not all of them; I am not sure if this is done by a select number of your writers, or some Wordpress plugin, but it is a good idea to keep these dates on every post URL, even just for the value of consistency
- Thinking about tags some more: your authors would ideally be taggable terms, binned under the category, People/Occupations/Authors
- I have done this for a handful of categories: *synonyms*; some terms are roughly synonymous, e.g., I say "toddlers" and a few of your writers tag with "lil kids"; since I have included an "alternativeLabel" property under the category Toddler with the value "lil kids", and since the trained GPT model can see this, it has this association; in any case, I bring this up because there are many more synonymous terms that can be added to the taxonomy

At your service,
TP

Semantic Science

- Ontology Engineering & Research Publication -

www.semanticscience.us