

Jennifer M. Martin. Crowd-based Information Organization: A Case Study of the Folk Ontologies in Wikipedia and TV Tropes. A Master's Paper for the M.S. in L.S degree. April, 2013. 68 pages. Advisor: Ryan Shaw

Folk ontologies are the organically created, crowd-managed ontology-like structures used to organize information in wikis. This paper takes as case studies the folk ontologies of the wikis Wikipedia and TV Tropes, performing a three-part analysis on them. The first part describes the folk ontologies and examines how they are maintained and used. The second part creates an evaluation framework based on the ontology evaluation literature and considers the folk ontologies in that light. The third part takes a random sample of pages from the wiki and performs a quantitative analysis looking at where and how the folk ontologies appear and what effects their presence has. The goal of the paper is to advance our understanding of how people, in aggregate, organize information.

Headings:

Wikis (Computer science)

Ontologies (Information retrieval)

Information organization

CROWD-BASED INFORMATION ORGANIZATION: A CASE STUDY OF THE
FOLK ONTOLOGIES IN WIKIPEDIA AND TV TROPES

by
Jennifer M. Martin

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

April 2013

Approved by

Ryan Shaw

Introduction

This paper provides an in-depth study of two “folk ontologies”. Folk ontologies are organically created, crowd-managed structures used in wikis to organize and make accessible the wiki’s contents. These structures provide a framework for understanding and navigating the contents of the wiki, giving a structured view of the knowledge contained in the wiki. The structures are similar in concept to the idea of a semantic wiki. A semantic wiki uses formal markup language to create machine-understandable semantic information about the contents of and relationships within the wiki. The main difference between folk ontologies and semantic wikis is that folk ontologies are informal, intended to be primarily human-readable, and designed to support browsing, whereas a semantic wiki structure is primarily formal, machine-readable, and designed to support querying. Also, folk ontologies generally do not have a strict hierarchy in their relationships, unlike semantic wikis which enforce is-a relationships between levels of the relationship structure. The folk ontology structures are generally built and maintained by the wiki members as part of the daily work on the wiki.

Folk ontologies are related to formal ontologies (traditional ontologies). Both types of ontology seek to define concepts, the relationships between those concepts, and instances of those concepts using a tree structure. Formal ontologies achieve this by capturing the information in one of the formal ontology languages, which are machine-readable and thus can be manipulated by computers. Formal ontologies are generally designed and maintained by specialists. Folk ontologies, on the other hand, accomplish

their objective using lists and hyperlinks. The structures of these ontologies are not inherently machine-readable¹ and thus cannot easily be manipulated by computers. Folk ontologies are designed and maintained by a given community, who are usually not ontology—or even information—specialists. Folk ontologies can also sometimes resemble taxonomies, as folk ontologies frequently include related-to relationships as well as is-a relationships.

Folk ontologies are also related to folksonomies (social tagging) in that both metadata structures are informal and crowd-created. The primary difference between the two is that folksonomies are flat, without any explicit relationships between terms, whereas folk ontologies are tree-based and designed to indicate relationships between concepts. Folk ontologies are generally also crowd-curated, with effort put towards consistency and non-redundancy, whereas folksonomies often have no curatorial oversight. While much research exists about folksonomies, folk ontologies have received no attention as such, even though they demonstrate that crowd-sourced metadata can be more sophisticated than mere clouds of keywords, as well as providing an example of how humans in aggregate naturally organize information. This paper aims to provide a basic understanding of how folk ontologies are structured, maintained, and used, which will hopefully provide a framework for understanding and discussing these metadata structures in the future, with a view towards furthering the understanding of how most people approach complex knowledge organization.

After a brief look at related literature, this paper takes as case studies the folk ontologies of the English-language Wikipedia (en.wikipedia.org) and TV Tropes (www.tvtropes.org). The examination will have three major parts: an explanation of the

ontology systems, a qualitative evaluation of the ontologies, and a quantitative analysis of how the ontologies are used.

Literature Review

Folk ontologies are not studied as such in the literature at all. The closest they come to being studied is when Wikipedia's folk ontologies are used to help create formal ontologies, but the focus on those papers is on extracting the information, not the qualities of the folk ontologies themselves. Folk ontologies are related to both the semantic web, particularly semantic wikis (as mentioned previously), and to collaborative formal ontology-building (as folk ontologies are built collaboratively), so a brief look at the work being done in all these fields is presented here as a means of grounding the discussion about folk ontologies in Wikipedia and TV Tropes.

Wikipedia has recently been recognized as a resource which can be mined to automatically create or refine formal ontologies; a number of projects have been undertaken to exploit this potential. Medelyan, et al, provide an overview of the research done in this area and related areas; they also provide an overview of the general structure of Wikipedia itself. While a number of different ontologies have been built based on Wikipedia and each uses a different approach in details, the generic process for extracting an ontology from Wikipedia is to define what is being extracted how, render the extracted information in a formal ontology language, and then evaluate and refine the resulting ontology. YAGO uses Wikipedia's navigational templates to mine data and refines it using Wikipedia's categories and the ontology WordNet (Suchanek, Kasnei, and Weikum). DBPedia mines Wikipedia's navigational templates (Auer, et al). Both

Ponzetto and Strube's ontology and KOG (Wu and Weld) derive from Wikipedia's ontologies.

Collaborative ontology building is an emerging field of study, covering a fairly broad range of types of research. The studies generally present a given software or structure and argue its fitness for collaborative ontology development; most of these are too specific to be relevant to folk ontologies, as the studies focus on specific ontology building environments (for instance, Palma, et al) or aspects of formal ontology building which do not apply to folk ontologies (for instance, Liu and Gruen). However, Maleewong, Anutariya and Wuwongse propose a "semantic argumentation approach" to support ontology building via a formal system of proposals, discussion, and iterative consensus building. While the specific semantic framework they propose is too formal to be used for folk ontologies, the general framework is applicable to ideas of how to structure the development of folk ontologies. Noy, et al, present a framework of different types of ontology building and the functional requirements for them; the types of ontology building can be applied to types of folk ontologies as well.

Semantic wikis, in theory, pull together collaborative ontology building with the ease of collaboration and editing found in wikis. Most of the literature reflects the quest to find a structure and editing format for semantic wikis which does not pose too high an entry barrier for the general web user, as that is a prerequisite for semantic wikis being widely adopted (see, for instance, Kuhn). Di Iorio, et al, have an interesting study, in that they examine tools which would allow the addition of semantic wiki properties to traditional wikis; this opens up interesting possibilities for potentially transforming folk ontologies into formal ontologies. Kousetti, Millard, and Howard discuss semantic wikis

in broad terms, looking at the nature of and requirements for creating quality semantic wikis. They also include a fairly comprehensive list of existing semantic wikis. Semantic wikis are interesting to consider relative to folk ontologies, because semantic wikis are primarily an information science-based push to codify and channel the behaviors that occur naturally in the creation of folk ontologies.

The Folk Ontologies

We turn now to our two case studies. Both Wikipedia and TV Tropes have three folk ontologies.² For each ontology, the following aspects are considered:

- What its purpose is within the wiki,
- What knowledge domain it draws from,
- Where and to what it is applied in the wiki,
- How and by whom it is applied,
- How and by whom it is managed and maintained,
- How the it interacts with the wiki at large, and
- How it relates to the other ontologies.

This section is designed to provide a detailed look at the actual functions and structures of the folk ontologies without making value judgments, both for learning purposes and to provide a background for the next sections.

Wikipedia

The English-language Wikipedia has three types of folk ontologies: categories, lists, and navigation templates (also known as infoboxes). All Wikipedia articles are supposed to be placed in at least one category;³ the other two types are optional. Any

article can have any combination of the folk ontologies associated with it; they are viewed as complementary organizational schemas.⁴

Categories form the backbone of knowledge organization on Wikipedia, as they are the only type of organization schema which is required. They group conceptually related pages together and structure the pages into a tree using subcategories. There are two sets of top-level categories—one with four main categories;⁵ the other with 24⁶—and all other categories can be accessed from those by drilling down through subcategories. However, these categories are not strictly hierarchical, as any given page or category can be a member of any number of categories. As Wikipedia is a general-purpose encyclopedia, and the categories cover all its articles, the categories are drawn from all knowledge domains. Categories are intended to be topical and objective, and arbitrary groupings are discouraged.⁷

Categories appear in two places: at the bottom of every article and on category pages. Thus, all categories and related pages are interlinked. Each article lists the categories it directly belongs to in a box at the very bottom of the page (see Figure 2). Category pages provide a brief introduction to what the category is about, or a link to a Wikipedia page with a more detailed description, and a listing of its subcategories and pages (see Figure 1). Categories have their own namespace, distinguishing them from the other types of pages. Only pages can be included in categories.

The categories can be created, maintained, and applied by all of Wikipedia's editors. Categories are added to pages by editing the page in question; new categories are created by adding the category name to a page as if it were any other category. The category page is automatically populated with all the pages and subcategories it contains;

the introductory section, which provides a brief overview of what the category is, can be edited by anyone. Any errors in categorization are handled by whoever notices them. Any major changes to (such as renaming) or reorganization of categories are supposed to be discussed first (in large part to prevent people from accidentally working at cross-purposes, and discussion must happen before categories can be deleted).

Category:Ice dancing

From Wikipedia, the free encyclopedia

The main article for this category is **Ice dancing**.



Subcategories

This category has only the following subcategory.

I

- ▶ Ice dancers (35 C)

Pages in category "Ice dancing"

The following 8 pages are in this category, out of 8 total. This list may not reflect recent changes ([learn more](#)).

- | | | |
|---|--|--|
| <ul style="list-style-type: none"> Ice dancing | F cont. <ul style="list-style-type: none"> Figure skating lifts Free dance (figure skating) | S <ul style="list-style-type: none"> Scala Eisrevue Short dance |
| C <ul style="list-style-type: none"> Compulsory dance | O | |

Figure 1: an example of a category page

V · T · E		Figure skating	[hide]
Disciplines		Single skating · Pair skating · Ice dancing · Synchronized skating	
Competition programs		Compulsory figures · Short program · Free skating · Compulsory dances · Original dance · Short dance · Free dance	
Elements	Required elements	Death spiral · Jumps · Lifts · Moves in the field · Spins · Spirals · Step sequence · Twist lifts	
	Jumps	Toe loop jump · Flip jump · Lutz jump · Salchow jump · Loop jump · Axel jump	
	Spins	Upright spin (Layback spin · Biellmann spin) · Sit spin · Camel spin	
	Moves in the field	3 turn · Bracket turn · Choctaw turn · Counter turn · Crossover · Mohawk turn · Rocker turn · Twizzle	
	Connecting elements	Besti squat · Cantilever · Hydroblading · Ina Bauer · Spread eagle · Split jumps · Walley jump	
Other		ISU Judging System · International figure skating · Adult figure skating · Four skating	
Historical interest		History of figure skating · 6.0 system · Special figures	
Lists		Figure skaters · Olympic medalists · Competitions · National championships · Terms	
		Category · Portal	

Categories (+⁺): [Figure skating elements](#) (-) (±) | [Pair skating](#) (-) (±) | [Ice dancing](#) (-) (±) | (+)

Figure 2: top: an example of a navbox; bottom: an example of a category listing on a page

Lists record pages that are part of a logical set, and can include links to sub-lists. Like categories, lists cover all knowledge domains; unlike categories, though, lists are not strictly topical nor do they have the restriction against arbitrary parameters, which is why lists are generally used for breaking groups of objects or people down by year, country, etc. They are frequently also used for grouping like types together.

Lists are created within the main namespace, and they are maintained and created in the same way as articles—anyone can create them, anyone can edit them, major changes should be discussed, and deletion must be discussed. List links are generally unidirectional—the list links to the pages in question, but generally the pages don't link back to the list.

Navigation templates come in two types, distinguished primarily by where they are placed in the article: sidebars go at the top right of articles (see Figure 4); navboxes go across the bottom of the article, just above the category listing (see Figure 2). These present related articles in an organized, consistent manner, making it easily visible to readers what information is related to the current article they are reading. Like the other two organizational schemas, navigation templates draw from the general sphere of knowledge; however, any given template draws from the specific subdomain of the articles it is associated with. A navigation template can be used on any page.

They are created and managed by consensus. Unlike lists and categories, they do not have individual pages as such, so they cannot be deleted in the same way; however, they can become deprecated and be removed from use. Navigation templates are generally used on all pages referred to in the template, so their links are generally bidirectional.

These are the three folk ontologies used in Wikipedia.

List of garden types

From Wikipedia, the free encyclopedia

A wide range of **garden types** exists. Below is a list of examples.

Contents [hide]
1 By country of origin
2 By historical empire
3 In religion
4 Other
5 See also
6 External links

By country of origin

[edit]

- Chinese gardens
- Dutch gardens
- English gardens
 - Anglo-Japanese style
 - Landscape garden
- French gardens
 - French formal garden
 - French landscape garden
 - Gardens of the French Renaissance



Figure 3: an example of a list

Lesser celandine

From Wikipedia, the free encyclopedia

Lesser celandine (*Ranunculus ficaria*, *syn.* *Ficaria grandiflora* Robert, *Ficaria verna* Huds.) is a low-growing, hairless **perennial plant**, with fleshy dark green, heart-shaped **leaves**. The plant is found throughout **Europe** and west **Asia** and is now introduced in **North America**. It prefers bare, damp ground and in the **UK** it is often a persistent garden weed. The **flowers** are orange, turning yellow as they age.

Ranunculus ficaria exists in both **diploid** (2n=16) and **tetraploid** (2n=32) forms which are very similar in appearance. However, the tetraploid type prefer more shady locations and frequently develops **bulbils** at the base of the stalk. These two variants are sometimes referred to as distinct sub-species, *R. ficaria ficaria* and *R. ficaria bulbifer* respectively.

According to the *Oxford English Dictionary*, *celandine* comes from the **Latin** *chelidonia*, meaning *swallow*. It was said that the flowers bloomed when the swallows returned and faded when they left.^[2] The name *Ranunculus* is **Late Latin** for "little frog," from *rana* "frog" and a diminutive ending. This probably refers to many species being found near water, like frogs.

Contents [hide]
1 Life cycle
2 In non-native locations
3 Medicinal uses
4 References in literature
5 See also
6 References
7 External links

Life cycle

[edit]

According to *Gilbert White*, a diarist writing around 1800 in the **Hampshire** village of **Selborne**, the plants came out on February 21, but it is more commonly reported to flower from March until May, and is sometimes called the "spring messenger" as a consequence.

Lesser celandine	
	
Scientific classification	
Kingdom:	Plantae
Division:	Magnoliophyta
Class:	Magnoliopsida
Order:	Ranunculales
Family:	Ranunculaceae
Genus:	<i>Ranunculus</i>
Species:	<i>R. ficaria</i>
Binomial name	
<i>Ranunculus ficaria</i>	

Figure 4: an example of a sidebar (right)

TV Tropes

TV Tropes is a site dedicated to identifying and recording the tropes (recurring patterns) that occur in all fictional media (not just TV). It has about 70,000 primary pages.⁸ It is a bit of an interesting case because the site is, as a whole, a type of folk ontology (or, more specifically, two highly overlapping folk ontologies): the ultimate goal is to define the tropes and document all instances of them in fiction, and to document all works and list all the tropes that appear in a given work. Thus there are trope pages with instances from works, and work pages with instances of tropes. Because of this, defining the boundaries between these folk ontologies and indices, the third folk ontology on the site, can get a little tricky.

Tropes are roughly hierarchical in nature, expressed in terms of supertropes (parent), subtropes (child), and sister tropes (related). These relationships are generally indicated in the body of the description of the trope, although there is a little-used relationship tool which makes explicit the relationships between tropes (see Figure 6). It is not a strict hierarchy, as a subtrope can have multiple supertropes. Trope descriptions can also have a “see also” area, which lists related tropes and sub- or supertropes, establishing a fairly rich web of references and relationships beyond the supertrope/subtrope tree and beyond mere hyperlinking.

Tropes are identified via observation of fictional works, and form their own knowledge domain. Articles for tropes are proposed by an individual and then must go through a community review and improvement process before they are launched to the main wiki. After that stage, any changes in definition must go through the Trope Repair Shop (TRS), which is a forum where proposed changes are discussed and decided on via

consensus. TRS also handles deletions. Anyone can add instances of the trope to work pages and instances where the trope shows up in works to the trope page; anyone can also delete any incorrect examples. If there is widespread misuse, a thread will often be started in one of the Projects forums to organize a concerted cleanup, performed by multiple editors.

Sharpened to a Single Atom

Weapons that have been sharpened to a single atom are a particular type of [Absurdly Sharp Blade](#) that are so sharp, their cutting edge can be measured in terms of atoms or molecules, often being just a single molecule wide. This usually gives the weapons [Absurd Cutting Power](#), but generally requires the blade to be made out of some sort of [Unobtainium](#) to maintain that sharpness without dulling instantly (or just shattering into a million pieces) the first time it's used. For these reasons, the trope is most frequently encountered in [Science Fiction](#) settings.

A common term used for this is "monomolecular". Most often seen with knives or swords, but occurs from time to time with [Razor Wire](#), too. This can be particularly troublesome, as monomolecular wire is usually functionally invisible, and walking into one can slice through vital bits [without being immediately noticed](#).

A weapon is only an example of this trope when a *specific* (very small) width is given for the blade's edge; if no specific measurement is given, then it's just an [Absurdly Sharp Blade](#).

Examples

OPEN/CLOSE ALL FOLDERS

ANIME AND MANGA

- Eve from *Black Cat* managed to manipulate her [Prehensile Hair](#) to do this later in the manga.
- In *Kiddy Grade*, Sinistra and Dextera's ship can split into two, creating a 'monodimensional' blade between the halves.
- Mamoru's katana from *Until Death Do Us Part* is a high tech monomolecular blade, able to break the bonds between molecules and cut anything (if the katana's angled properly). At one point he cuts a gun in half; when the pieces are pushed back together, it appears as if the gun was undamaged, because the cut is that sharp.

COMICS

- Nemesis, a member (and former enemy) of *Alpha Flight*, uses a saber whose blade is only a single molecule wide.
- In *Kingdom Come Wonder Woman* has a sword that is sharp enough to "carve the electrons off an atom". Leaving aside the ways that doesn't actually

Figure 5: an example of a trope page

Absurdly Sharp Blade has these relationships to other tropes:

parents	kids	shares a parent with:	
Absurd Cutting Power	Sharpened To A Single Atom	parent Absurd Cutting Power	child Vibroweapon
		"	Razor Floss
		"	Laser Blade
		"	Hot Blade
add a parent	add a kid		
Stick to defining <i>direct</i> parental relationships. This isn't a good place to lump grandparents and great grandparents all into one list.	See the note in the parent column about direct relationships. Also, adding a parent to a kid is the same action as adding a kid to a parent. It is not necessary to do both.		
		<input type="button" value="add"/>	

Figure 6: an example of the relationship tool

Works form a very simple folk ontology that mimics the way they are present in the real world: franchise pages link to the works which comprise the franchise and series pages link to the individual works in the series (unless the series is treated as a single entity). Like trope pages, those work pages will be interlinked, so one can navigate in either direction through the hierarchy.

Works pages refer to conceptual and physical entities (books, films, TV shows, etc.) and thus have a one-to-one correspondence with the general class of fictional works. Anyone can create a works page at any time,⁹ and anyone can expand upon or correct errors within a works page. Any sort of large-scale reorganization of the works folk ontology (such as changing the way a franchise page is laid out) is recommended to be discussed first.

Indices organize, categorize, and provide browsing access to both trope and work pages. Each index generally covers either a topic or a genre, and they have a loose tree structure similar to tropes. In fact, in a way, they are part of the same tree as the tropes, even though they are conceptually different; indices bleed into tropes, in that trope pages are listed at almost every level of index, and high-level supertropes can also function as indices. The main difference between an index and a high-level supertrope is that the index is a collection of loosely related tropes (for example, all tropes about warfare) whereas a high-level supertrope is a collection of tropes which are a subtype of the supertrope. Another way of looking at it is that supertropes/subtropes are true hierarchies in that every subtrope is an instance of the supertrope, whereas indices are trees of related objects.

Franchise: Harry Potter

Harry Potter began as a series of seven fantasy novels by J.K. Rowling and later expanded into a multimedia franchise. We have separate pages for its book, film, and video game incarnations; this one is a disambiguation page to help you locate the more specific one you were looking for. Note that there is also a *Harry Potter card game*, unlisted and without a page but noted in the indexes.



- *Harry Potter*, the 1997-2007 book series consisting of 7 novels and 3 spinoffs:
 - *Harry Potter and the Philosopher's Stone*
 - *Harry Potter and the Chamber of Secrets*
 - *Harry Potter and the Prisoner of Azkaban*
 - *Harry Potter and the Goblet of Fire*
 - *Harry Potter and the Order of the Phoenix*
 - *Harry Potter and the Half-Blood Prince*
 - *Harry Potter and the Deathly Hallows*
 - *Fantastic Beasts and Where to Find Them*
 - *Quidditch Through the Ages*
 - *The Tales of Beedle the Bard*
- *Harry Potter*, the 2001-2011 series of film adaptations on the aforementioned books.
- *Harry Potter*, a series of video games, most based on straight adaptations of the films, although including some original titles.
 - *LEGO Harry Potter*, a duology of LEGO Adaptation Games.

Also see *A Very Potter Musical* and its sequel, a popular set of fan musicals; *Potter Puppet Pals*, another famous fan production; an up-and-coming fan production *Harry Potter And The Ten Years Later* which takes Harry into the real world; and the notoriously bad fanfic *My Immortal*.

Kajjudo	Card Games	Legend of the Five Rings
The Collector	Muggle Quidditch	The Illustrated Guide to the Order

Figure 7: an example of a franchise page

Literature: Harry Potter and the Chamber of Secrets

"It is our choices, Harry, that show what we truly are, far more than our abilities."
— Albus Dumbledore

Second book in the *Harry Potter* series. Published in 1998. The only parts which seem to have been absorbed by Popcultural Osmosis are Dobby warning Harry not to go to Hogwarts this year and that Flying Car. After all, they were prominently in the trailer for The Film of the Book.

The main plot involves the Chamber of Secrets, a hidden chamber within Hogwarts built by Salazar Slytherin. A big fan of Fantastic Racism, Slytherin built the Chamber to house a monster which can only be controlled by his heir and which is intended to attack all those Muggle-borns "unworthy to study magic". Now, someone has opened the Chamber, implying the Heir of Slytherin has returned to Hogwarts, but who is it?

You may have noticed this storyline has rather little to do with the overall Story Arc. *Chamber* is often accused of essentially being devoted to a Wacky Wayside Tribe for this reason. In reality, the book is just an Innocuously Important Episode, and introduces a major Chekhov's Gun among other bits of Foreshadowing for several later books, particularly *Harry Potter and the Half-Blood Prince*.

Tropes exclusive to this book or at least especially prominent in it:

- Absurdly Spacious Sewer: What kind of building has pipes big enough for a to move around in?
- Are You Sure You Can Drive This Thing?: "No, but we're far too anxious to consider another plan."
- And I Must Scream: The petrified victims of Tom Riddle's Basilisk.
- Bathroom Stall of Overheard Insults
- Becoming the Boast: Subverted. Lockhart seems like a Know-Nothing Know-It-All with a grossly inflated ego. In reality, he's exactly that, also!

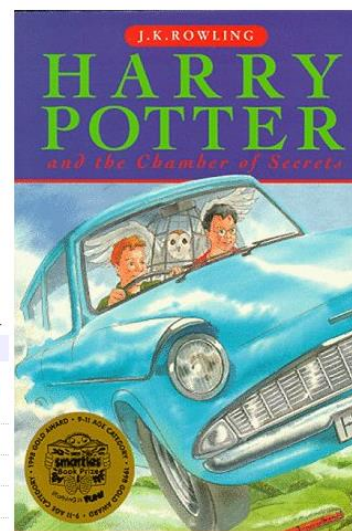


Figure 8: an example of an individual work page

As an example of this intermixing, consider the pages “[Motifs](#)” and “[Dialogue](#).” Motifs is a high-level supertrope and Dialogue is a top-level index. Both pages contain a list of tropes which fall into them, and both lists use the wiki’s index functionality (which causes the index page to appear in the index listing at the bottom of the pages in the index). However, Motifs only contains tropes which *are* types of motifs, whereas Dialogue contains tropes which are *about* or *related to* dialogue (for instance, “[Everyone Gets Their Turn](#),” which is about the tendency of television shows to apportion speaking time evenly among all participants in a conversation) in addition to tropes which are forms of dialogue. Dialogue has both a list of tropes and a list of sub-indices, such as [Accent Tropes](#).” Accent Tropes has its own list of tropes that fall under it, and could have sub-indices, which would have their own list of tropes. For a second example, consider the trope “[Accidental Pun](#).” The bottom of the page lists the indexes it is on (see Figure 9 for an example of such an index listing). Using the index listings to navigate can take us from “Accidental Pun” to “[Pun](#)” (subtrope to supertrope), from “Pun” to “[Double Meaning](#)” (subtrope to supertrope), and from “Double Meaning” to “Dialogue” (trope to index), all while using the same navigational mechanism. In this way, the wiki’s navigational features blur together indexes and tropes, despite the conceptual distinction between the two.

Royal Rapier	I Like Swords	Sinister Scimitar
Sentient Phlebotinum	Applied Phlebotinum	Shoe Phone
Absurdly Sharp Blade	Combat Tropes	Acrofatic
Absurdly Sharp Blade	Absurd Cutting Power	Hot Blade
Serial Escalation	Rule of Cool	Shoulder-Sized Dragon
Absurdly Sharp Blade	Weapons and Wielding Tropes	An Axe to Grind

Figure 9: an example of a listing of the indices a page belongs to, found at the bottom of a page (this is from “Sharpened to a Single Atom”; see Figure 5). Each box indicates a single index; the name of the index is in the center (in black). On the left and right are links to the previous and next pages in the index, respectively.

Distraction Tropes

Tropes that cover things that distract or divert a character's attention, whether by accident or design, as well as the characters that plan or fall for them.

For examples of things that might distract you, the reader, from something you really should be doing instead, see Wiki Walk and TV Tropes Will Ruin Your Life.

Tropes:

- Attention Deficit... Ooh, Shiny!
- Cacophony Cover Up
- Centipede's Dilemma
- Convenient Decoy Cat
- Crashing Through the Harem
- Delicious Distraction
- Did You Just Flip Off Cthulhu?
- Distracted by the Sexy
 - Distracted by My Own Sexy
 - Not Distracted by the Sexy
- Distracting Disambiguation
- Easily-Distracted Referee
- Fake-Out Make-Out



They have their uses

Figure 10: an example of an index

Pun

A pun is a form of word play where a word with more than one meaning is exploited to make a joke based on this double meaning. Usually done for humorous effect.

For trope names that are puns go to Just for Pun. For tropes that are pun names of other tropes, you want Snowclones. These are very PUNNY.

Associated Tropes

- Accidental Pun
- Double Entendre
- Duck!
- Epunymous Title
- Feghoot
- Flintstone Theming
- Goroawase Number
- Hurricane of Puns
- Lamé Pun Reaction
- Multiple Reference Pun
- Parallel Porn Titles
- Pun-Based Title



Now is the winter of our disco tent.

Categories:

- Just for Pun
- Snowclones

Figure 11: an example of a high-level supertrope that also functions as an index, for comparison

Work indices are much more clearly defined, as they are lists of works by genre, and it is quite easy to tell the difference between a list of works and a page about a specific work or set of works.

Indices are generally managed in much the same way as trope pages, although the tropes tend to get more attention. Every page is supposed to be on at least one index, and the indices a page is on are listed at the bottom of the page.

These are the three folk ontologies used on TV Tropes.

Comparison Between the Two Wikis

Although Wikipedia is a general purpose encyclopedic wiki whereas TV Tropes covers a specialized domain, the wikis' folk ontologies and manifestations of the folk ontologies do have certain similarities. This section explores Wikipedia's categories and lists relative to TV Tropes' work/trope/index complex and Wikipedia's navigational templates relative to TV Tropes' see also sections and relationship tool.

TV Tropes' work/trope/index complex of folk ontologies forms the backbone for organizing information on the site, in much the same way that Wikipedia's categories form the backbone for organizing the information on that site. Both those (sets of) folk ontologies seek to define the entire domain—whether it's fiction and the tropes used in fiction or all knowledge—the site encompasses.

In terms of manifestations of the folk ontologies—that is, the way the folk ontologies are recorded within the pages themselves, as opposed to the conceptual version which exists purely as a logical concept—TV Tropes' indexes and high-level supertropes are rather similar to Wikipedia's categories and lists. A single page for either defines a single concept, which can be topical, geographical, time-related, etc., and then

lays out which other concepts (pages) are encompassed within that concept. This creates a tree of access, although the fact that most pages are entered in multiple indexes, categories, etc., generally means that there is no single way to get a given page. The pages form more of a web than a true tree.

Wikipedia's navigational templates have similarities to TV Tropes' see also sections and relationship tool. Both the navigational templates and the see also sections have certain similarities to taxonomies, in that they point readers to broader, narrower, and laterally related articles. However, both forms draw from the underlying folk ontology of the site to give those related articles, and thus they are both manifestations of their respective ontologies. The navigational templates arrange the related articles into a well-defined, dense form, which TV Tropes' relationship tool also does. All three provide "at a glance" information about how a given page relates to other pages on the site.

These similarities highlight the fact that both wikis use the folk ontologies as a primary means of organizing and providing navigation for their site.

Qualitative Analysis

This section is intended to give an idea of how well folk ontologies work as ontologies by performing a qualitative evaluation of the folk ontologies. The framework for this is drawn from the professional literature on the evaluation of formal ontologies and presented below. The folk ontologies of each wiki are then evaluated in terms of the framework.

The Framework

One of the biggest challenges for building a framework for evaluating folk ontologies is that folk ontologies are not defined using a formal, machine-readable

language. Most ontology evaluation literature assumes, quite reasonably, that the ontology under consideration is in a machine-readable format and thus can be automatically processed in a variety of ways (see, for instance, Solskinnsbakk, et al, and Navigli, et al), and does not pay much attention to the underlying concepts. However, because folk ontologies are not defined in machine-readable language, they are not amenable to any sort of automatic processing, and thus can only be evaluated via the underlying concepts. As a result, the framework developed for this paper looks at the general concepts underlying “quality ontologies” and considers how to apply those concepts in an evaluation of the folk ontologies.

One major difference—and limitation—of this evaluation of the folk ontologies is that it is impossible to actually evaluate the entire ontology. When a machine-readable ontology is processed using automatic evaluation techniques, the entire ontology can be evaluated. However, as automatic techniques are not available for evaluation of the folk ontologies and as they run into thousands of terms, the folk ontologies cannot be exhaustively examined. The evaluation presented in this paper will draw from the author’s general familiarity with the ontologies, obtained by exploration and use of them, as well as from spot checks of portions of the folk ontologies.

Five primary concepts for measuring the quality of ontologies appear broadly in the literature. These five primary concepts are context, consistency, completeness, conciseness, and flexibility. These five concepts form the core of the evaluative framework for the folk ontologies.

An ontology exists in a context, and the context must be considered when evaluating the quality of an ontology (Pak and Zhou 14; Sabou and Fernandez 197;

Vrandečić 297). This includes considerations of whether the ontology is a good fit for the purposes it is intended to be used for (Pak and Zhou 14; Sabou and Fernandez 199; Vrandečić 296) and of how well it interacts with any other ontologies, applications, etc., in the same context (Pak and Zhou 14; Sabou and Fernandez 199; Vrandečić 297). To apply this to a folk ontology, the folk ontology's role on the site should be considered: how does the site use the ontology? Is it robust enough for that purpose? Does it have a sufficient technical support for that purpose? What other aspects of the site does it interact with? Does it mesh well? Is there more than one folk ontology? If so, how well do they play together?

A consistent ontology is one that does not allow contradictions (Gómez-Pérez 394; Pak and Zhou 15; Sabou and Fernandez 197-98; Vrandečić 296). This includes verifying that the definitions do not contradict themselves or the real world (metaphysical consistency), that the formal definition of the concept matches the informal understanding of the concept (internal consistency), and that no contradictory statements or conclusions can be derived from the structure of the concepts (inferential consistency). To apply this to folk ontologies, a conceptual analysis of the components of the folk ontologies can be performed: do the definitions of the concepts make logical sense? Are the terms used in a way that is consistent with their definition? Are any pages placed in categories which are (or logically should be) mutually exclusive?

A complete ontology is one that contains all relevant information (Gómez-Pérez 394-95; Pak and Zhou 15; Sabou and Fernandez 198; Vrandečić 296-97). This includes verifying that the ontology is not missing any information within the definitions and is also not missing any definitions on the whole. It also means ensuring that the definitions

are clear and understandable. For all that completeness is one of the major considerations of ontology quality, it is impossible to actually prove completeness (Gómez-Pérez 394); the most an evaluation can prove is a lack of obvious incompleteness. To apply this to folk ontologies, the concepts, definitions, and pages must be considered relative to the knowledge domain they are seeking to encapsulate: are there any concepts which you would expect to find in the folk ontology but cannot? Do the definitions cover all the most pertinent bits of information? Are the definitions easy to understand? If you are drilling down through the folk ontology, are there any particularly large jumps from general concepts to very specific ones (suggesting a lack of intermediate concepts)? Are there any terms which seem only laterally relevant to where they are placed in the folk ontology, yet there does not appear to be any more fitting place for them?

A concise ontology is one that contains only the needed definitions and relationships (Gómez-Pérez 395; Pak and Zhou 15; Sabou and Fernandez 198; Vrandečić 296). This includes removing any irrelevant definitions, ensuring that the concepts are precisely defined, and managing redundancies. Conciseness does not necessarily demand that all redundancy be removed; controlled redundancy can be used to help identify and clarify definitions of concepts (Pak and Zhou 15). However, no explicit redundancies should exist (i.e., definitions should not state that they are the same as something else) and redundancies should not be easily inferred from other concepts or relationship in the ontology (Gómez-Pérez 395; Pak and Zhou 15). To apply this to folk ontologies, the definitions of the concepts should be checked: are there any pages which are essentially duplicates? Do the definitions clearly lay out the boundaries of the concept they cover? Are all the concepts present actually useful for understanding the knowledge domain?

A flexible ontology is one which can handle additions and small changes gracefully (Gómez-Pérez 395; Vrandečić 295). This includes the ability to add new concepts to the ontology as needed without having to remove relationships or significantly alter existing definitions and the ability to make small changes in definitions without destroying the integrity of the ontology. To apply this to a folk ontology, the methods for adding new concepts to the folk ontology and the methods for editing existing definitions should be considered: how smoothly are new concepts added? How much leeway for minor definition changes is there? How are any other, more major changes handled?

The following section applies this evaluation framework to the folk ontologies of Wikipedia and TV Tropes.

Evaluations

Wikipedia

Context: Wikipedia's folk ontologies are used primarily to support browsing. The category structure is very extensive and provides lots of access points. Lists and navigation templates cover more narrowly defined knowledge domains, seeking to encapsulate a complete overview of their domain. Navigation templates were designed specifically to encourage browsing between the articles in them, and thus have the best format for that purpose. Categories are supported by special software which creates the listing on the category pages automatically, easing maintenance; however the links are buried at the very bottom of the articles. Lists are the weakest navigational aid because they only link in one direction; they don't use any technology beyond the basic linking functionalities of the wiki. All three types of folk ontologies interact can interact with all

the pages on the site, especially article pages. The technical process of adding them to pages is no more difficult than adding any other sort of content to a wiki page, although adding navigational templates requires an understanding of the template structure, which can be quite complex. The three folk ontologies supplement each other. Categories provide the overarching backbone; navigational templates pull out specific subdomains and treat them in more detail, making the relationships more obvious on the article pages; and lists cover certain topical and conceptual groupings which are generally not included in the categories.

Consistency: Wikipedia has no specific method to prevent inconsistency aside from the attention of its editors. That said, the editors are encouraged to make any changes necessary to keep the folk ontologies in good shape, and that includes fixing any inconsistencies. Wikipedia categories are generally metaphysically consistent (they accurately reflect the real world) and internally consistent (the general understanding of what the category means matches the formal definition of the category). However, they lack inferential consistency; drilling down into subcategories usually reveals (several layers deep) subcategories which do not logically belong in the parent category, and sometimes even outright logically contradict the parent category. An example of the latter is the category [Artificial satellites orbiting Earth](#) being found in the [Nature](#) category via the category [Earth](#). This is a result of Wikipedia's categories not requiring strict is-a relationships; related-to relationships are also permitted.

Completeness: Wikipedia operates on the assumption that its folk ontologies are incomplete, which is why their policy pages for them encourages editors to create any new instances of one of the folk ontologies that is felt to be missing. That said, the

existing structures are quite extensive and reflect a broad spectrum of the human knowledge. The top level concepts are well-represented and do not appear to have any major gaps, so the incompleteness really only enters in when you drill down into the more specific areas of the folk ontologies. The categories do have an issue with completeness, though, beyond the generic incompleteness of not having yet captured the full spectrum of human knowledge: they are often lacking definitions. This means that the meaning of the category must be deduced from the name along; often, this is adequate, but the folk ontology would become more robust if written definitions were used more broadly.

Conciseness: Conciseness is rather difficult to define for a folk ontology that aspires to capture all notable knowledge. Minor issues of conciseness, such as wordiness, can be handled by whatever editor happens to notice the problem. More major issues of conciseness, such as whether a particular instance of one of the folk ontologies is needed, are referred to the community at large and decided through discussion. Cases of duplication are solved by merging the instances in question.

Flexibility: This is an area where folk ontologies excel. As mentioned previously, editors are encouraged to add new concepts and definitions as they see the need, and the ontologies are designed to allow easy addition of new concepts. Anyone can add relationships between existing concepts, and if a new concept would slot neatly into an empty space without requiring rearrangement, it can simply be added. If a new category would necessitate the rearranging of existing structure, discussion among the community is needed. Also, the relationships between the concepts can be rearranged if it becomes clear that the existing structure is less than ideal. Of course, the downside to this much

flexibility is that the folk ontologies are constantly in flux; there really is no single, stable state for most of them.

TV Tropes

Context: TV Tropes' folk ontologies exist to support both browsing and finding. All three are extensive, and they all draw off the same technology, which is sufficient for the purpose. The three folk ontologies together form the basis for the site. As mentioned in the description of the folk ontologies, the index and trope folk ontologies, while conceptually distinct, blend together in practice. For this reason, they will be primarily considered as a whole in the following evaluation.

Consistency: The works folk ontology is quite consistent, aside from the occasional confusion due to multiple works having the same name. However, the tropes folk ontology struggles with consistency on an ongoing basis; quite frequently, a trope will have a formal definition but will be used to mean something else. These cases are handled in the Trope Repair Shop, with the goal being to gain consistency between usage and formal definition. Inconsistency within a definition or through conflicting indexes is much rarer and can be generally solved by simply removing the part that is wrong, as such issues are rarely actual paradoxes. While definitional inconsistency is rare, inconsistency in assigning instances is quite common, even when the definition is generally well-understood; this is dealt with by whoever notices deleting the misplaced instance.

Completeness: Neither the works nor the trope folk ontologies are complete. The wiki has approximately 35,000 work pages, which is a tiny fraction of all the existing fictional works in existence. The wiki also operates on the assumption that there are

many more tropes still to be recognized and defined. Aside from this basic assumption of incompleteness, another issue with the trope folk ontology is a lack of higher level tropes being actually defined; often, the more specific tropes will be defined but the overarching concept will not exist in the folk ontology.¹⁰ Another common struggle with completeness on the wiki is confusing or unclear definitions; this issue is also addressed in the Trope Repair Shop, where the editors will discuss and come to a consensus on what the definition ought to be and rewrites the description to be clearer.

Conciseness: Conciseness is rarely a problem for the works folk ontology; occasionally duplicate pages are created, but otherwise, the goal is to have a page for every work in existence, so there is considered to be no such thing as an unneeded work. Tropes and indexes, on the other hand, can be deemed too specific or meaningless, at which point they will be removed from the wiki (and thus from the folk ontology). Duplicates are also an issue with tropes; when noticed, the issue is solved by merging the tropes together. A certain amount of redundancy in tropes is tolerated to allow fine distinctions, though.

Flexibility: Both folk ontologies are very flexible and have new entries created daily. Works can be created by anyone, and are simple to insert into the works folk ontology, as it is a fairly flat ontology. Tropes go through a draft process before being formally introduced into the trope folk ontology; this helps weed out bad concepts, duplicates, poor definitions, etc., and helps gather instances and establish where it will fit into the folk ontology. While minor shifts in the trope folk ontology are quite common, such as adding new relationships or inserting a supertrope, large-scale reorganization of

the structures is quite rare and is quite difficult to do, given the complex web that the trope folk ontology is.

Conclusions

For both Wikipedia and TV Tropes, the folk ontologies are quite flexible at the cost of stability, which makes consistency, completeness and conciseness hard to measure. Also, because the folk ontologies are constantly growing and being modified, the refinement process is ongoing, meaning that at any given time there are almost guaranteed to be inconsistencies and inconcision. Formal ontologies avoid that by having a stable form, which is refined to remove those issues before being put in use. Both wikis also operate on the assumption that their folk ontologies are incomplete, TV Tropes more strongly so than Wikipedia.

TV Tropes' trope folk ontology has the poorest quality out of all the folk ontologies considered, with its rampant problems with consistency, completeness, and conciseness. It is probable that this is largely due to the fact that the folk ontology is defining the knowledge domain it occupies as it is developed and is an end in itself: rather than organizing other knowledge or material, it is organizing itself. The works folk ontology and Wikipedia's folk ontologies have the advantage of referring to already created and conceptualized knowledge domains which refer to physical objects and concepts outside of the ontologies themselves; in contrast, TV Tropes is the first large-scale project to identify the tropes that occur in works and thus its folk ontology *is* the knowledge domain. This means that it is dealing not only with issues of conceptualizing a knowledge domain into an organized schema, but also with concurrently delimiting and

defining the knowledge domain in and of itself, which is why there are so frequently problems particularly with inconsistency, poor definitions, and lack of concision.

Quantitative Analysis

This section is designed to give a more concrete, detailed view of how the folk ontologies are used within the wikis and to suggest the ways in which the folk ontologies affect the use of the wikis. This is accomplished via analysis of a random sampling of fifty pages from each wiki. For each page, the degree of presence of the folk ontologies was recorded, as well as a variety of other metrics about the page, such as creation date and number of edits. Conclusions are drawn from trends observed in the data.

The following subsections for each wiki look at the methodology used to collect the data, report on the qualities of the sample of pages, and examine trends noticed in the data. Two-tailed Pearson's r correlations were used to determine statistical significance. All scatterplots created from the data can be seen in Appendix A; the following section only displays the most relevant ones. A final subsection compares the results between the two wikis.

Wikipedia

Methodology

The primary means of sampling was Wikipedia's "Random article" link; however, that link only returns article pages and not category pages, the latter of which were also wanted in the final sample. Therefore, 45 articles were collected from the random article link. The remaining five articles were category pages. These were collected by going to the first category of the first random article, the second category of the second random

article, and so on, up to five. The inclusion of lists, which are a type of article page, was left up to chance.

For each page, the number of categories the page belonged to, its creation date, when it was last modified, its page length, the number of redirects to the page, the number of other pages linking to that page (wicks), the total number edits, and the total page views in the last 30 days were recorded. The wicks were recorded in units of twenty, to simplify counting, as Wikipedia does not display an exact count. The data were collected on March 13, 2013, so all values reflect the state of the pages as of that date. For article pages, the number of sidebars and the number of navboxes were also recorded; category pages do not use either of those forms of folk ontologies. For category pages, the number of subcategories and the number of pages within the category were also recorded.

General Results

The final sample contained 50 pages, of which 45 were article pages and 5 were category pages. Of the article pages, 23 were healthy articles, 15 were stubs, 3 were disambiguation pages, 2 were orphaned stubs (meaning no other articles linked to that page), 1 was a list, and 1 was an orphaned non-stub article. Categories were the most consistently used type of ontology, with a 100% use rate and an average of just under four categories per page. Sidebars were used on about two-thirds of the article pages, and navboxes were on less than a third. For categories, three of the five had subcategories, and all contained pages. The category sample size is too small to say much about anything category specific. Table 1 summarizes the captured data, showing the mean, median, mode, minimum, and maximum for each aspect recorded.

	sidebars (articles only)	navboxes (articles only)	categories	last modified	page length (bytes)
mean	0.622	0.378	3.92	11/3/2012	5201.5
median	1	0	3	2/15/2013	2997.5
mode	1	0	2	2/27/2013	no mode
minimum	0	0	1	11/10/2010	45
maximum	2	2	12	3/13/2013	45017

	creation date	number of edits	wicks	page views	redirects
mean	2/8/2008	53.38	42	1480.32	0.84
median	6/19/2007	21.5	20	259	1
mode	no mode	16	20	85	0
minimum	2/2/2002	2	0	18	0
maximum	2/4/2013	283	260	30718	5

	subcategories (categories only)	contained pages (categories only)
mean	9.8	1735
median	1	239
mode	0	no mode
minimum	0	3
maximum	43	7983

Table 1: Wikipedia data summary

Observations

One of the first things taken into account was whether or not the usage of one form of folk ontology (sidebar, navbox, category) correlated with usage of any of the other forms of folk ontology. As can be seen in Figure 12, only sidebars and categories show any clear pattern: any page with more than five categories has at least one sidebar. Thus the presence of increasing numbers of categories correlates with the presence of sidebars ($r=0.491$, $p<0.05$). Looking at the graph, the presence of large numbers of categories is predictive of at least one sidebar, but the presence of a sidebar does not lend itself to predicting the number of categories. Navboxes showed no significant correlation

with either of the other two forms of ontology ($r=0.093$, $p>0.05$ for navboxes and sidebars, and $r=0.173$, $p>0.05$ for categories and navboxes).

One unexpected discovery was that the greater presence of sidebars or categories cut off the “long tail” for when an article was last edited (for sidebars, $r=0.312$, $p>0.05$; for categories, $r=0.284$, $p>0.05$; see Figure 13). If an article page had at least one sidebar or four categories, then it had been edited within the past year. Pages with fewer instances of the folk ontologies did not have that guarantee. This suggests either that folk ontologies on Wikipedia are beneficial for attracting editor attention or that pages which receive greater editor attention are more likely to be given sidebars and categories.

Another upward trend occurred with page views: more categories correlated with greater numbers of page views ($r=0.515$, $p<0.05$). The plot for the categories shows the lift in the spread of data points as the number of categories goes up (see Figure 14), but it is not as strong as that for sidebars. This suggests that readers of Wikipedia may use categories to find pages, and thus more categories make the pages more visible. Alternatively, it may be that more page views lead to more categories being added, given that pages which get the most views also receive the most edits ($r=0.682$, $p<0.05$). It is possible that part of the reason sidebars and navboxes are used less is because there are fewer of them overall, but it could also be that categories see more use because they are simpler to add to a page, as they don’t require templates.

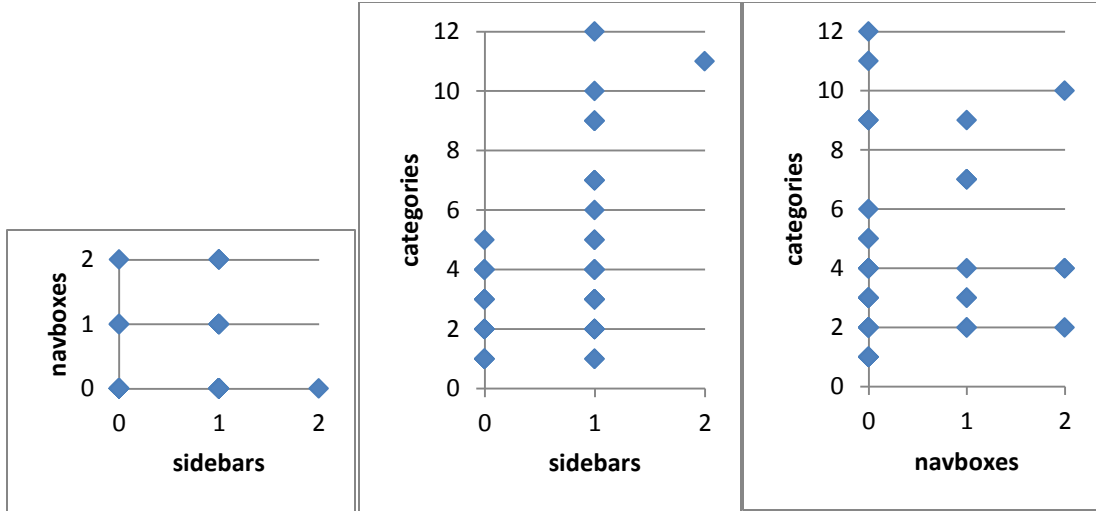


Figure 12: the usage of the types of folk ontologies relative to each other

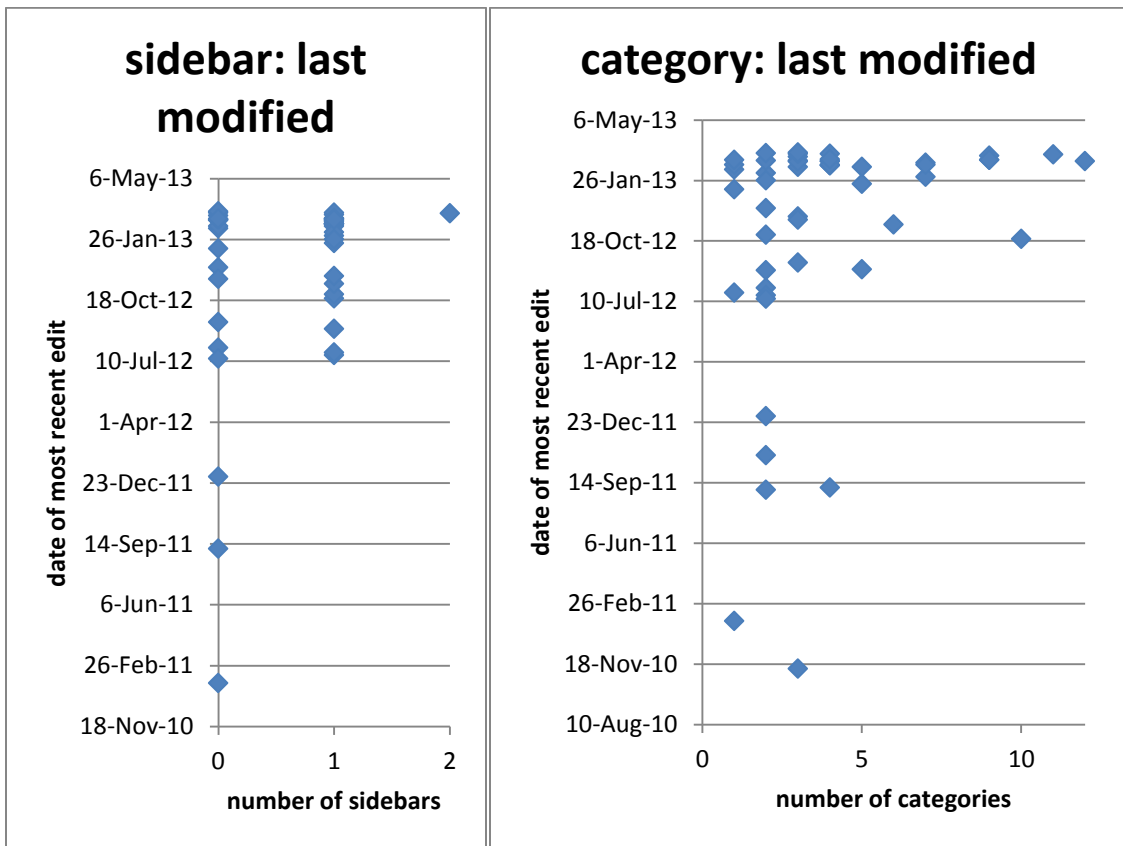


Figure 13: when a page was last modified relative to the number of instances of folk ontologies

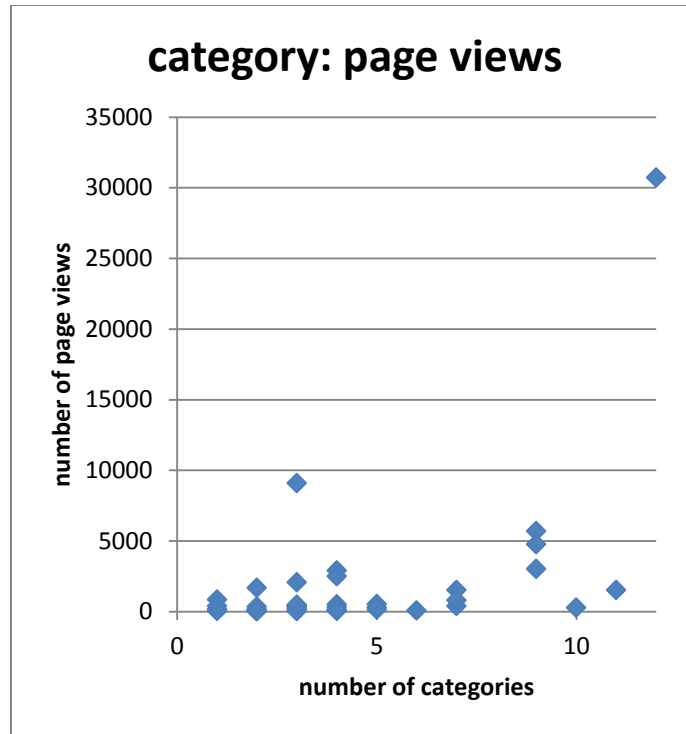


Figure 14: the number of page views relative to the number of categories on the page

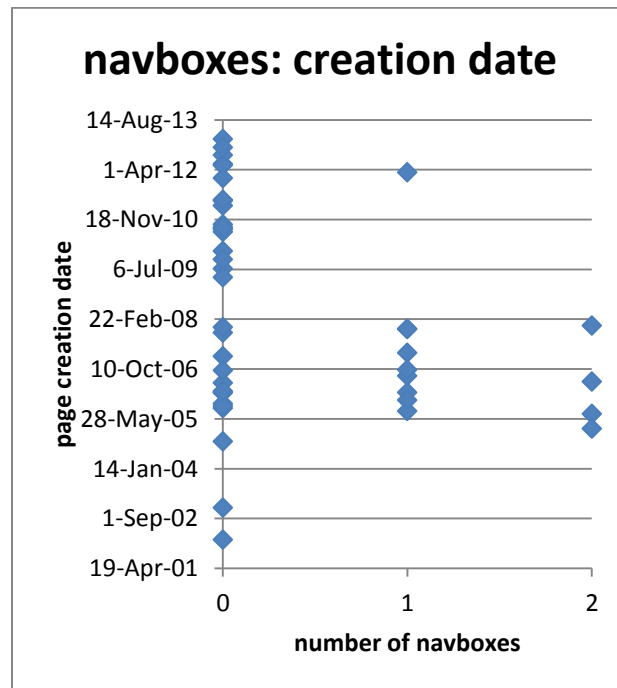


Figure 15: the creation date versus the number of navboxes

Possibly related to the relative rarity of navboxes, they are most likely to appear on articles which were created in 2004-2008 (see Figure 15). There was only one article with a navbox which was created more recently than February 2008, despite the fact that a large number of the articles in the sample were created past that date. The correlation between navboxes and article creation date is $r=-0.308$ ($p<0.05$). This suggests that navboxes may have fallen out of favor with the editors of Wikipedia (a greater sample would need to be considered for a definite conclusion), even though navboxes are certainly not deprecated in Wikipedia's documentation. If it is true that navboxes have fallen out of common use, then that would explain why navboxes only appear on one-third of articles: navboxes were mostly added more than five years ago, so almost all articles created past that date—which is quite a lot—do not have navboxes.

All-in-all, categories appear to be the most strongly correlated with measures of attention, such as number of edits to a page ($r=0.476$, $p<0.05$), last edited date, number of page views, and number of wicks ($r=0.317$, $p<0.05$). Sidebars and navboxes have few statistically significant correlations with any of those measures. These correlations suggest that either that categories help attract attention to pages or that pages which receive attention have more categories added to them (which in turn would suggest that the editors of Wikipedia consider a greater number of categories to be an improvement); further research would need to be performed to determine the direction of causation.

TV Tropes

Sampling Method

Sampling for TV Tropes was somewhat complicated due to the stratification of certain types of pages. Like with the Wikipedia sample, the primary means of sampling

was from TV Trope's "Random" button. However, that button only returns trope pages which are not marked subjective (subjective tropes include audience reactions and tropes which on the surface would appear objective but in practice engender a lot of disagreement over what fits). In the interests of looking at all the primary pages on the site, work pages, index pages, and subjective trope pages also needed to be included, which required other means of sampling.

Thus, 30 trope pages were collected via the random button. From those pages (plus several more random trope pages), ten work pages were collected: the tenth work from the first folder of the first random page, the tenth work from the second folder of the second random page, etc.; if a page did not have at least ten works in the selected folder (which was true for over two-thirds of the pages), no work was chosen from the page. Five indices were chosen from the list of all indices in the main namespace;¹¹ the indices X number from the top of each of the three columns and from the bottom of the two side columns, with X supplied by rolls of a twenty-sided die. Pages which were also trope pages were ignored and the dice rerolled for a further count, as the purpose was to select pure indices. Five subjective trope pages were selected; two from the top and one from the bottom of the audience reaction index,¹² and one from the top and one from the bottom from the index for other subjective tropes.¹³ The distance from top or bottom was once again determined by rolling a twenty-sided die. Thus, as random a sample as possible including all the main types of pages was selected for TV Tropes.

For each page, the number of indexes it was on, its length, its creation date, the number of times it had been edited, its number of wicks, the number of redirects to the page, and its inbound count were recorded. The data were collected on March 13, 2013,

so all values reflect the state of the pages as of that date. The inbound count is the number of hits the page has received from non-search-engine external links since the beginning of 2012; TV Tropes does not record a generic page view count. Date of creation is an approximation, as TV Tropes does not record that information. This estimate was obtained from the earliest date of the following: the date of the first edit in the edit history (the edit creating the page is never preserved in the page history, and the edit history was purged in August 2010, so current histories only go back to then); that the page was created before the current version of discussion was installed, which happened around June 2010 and is indicated by the presence of archived discussion; or the date of the last edit on the trope's draft in the trope page creation system. For trope and index pages, the number of tropes indicated in the "see also" statement was recorded; work pages do not have "see also" statements. For trope pages, the number of supertropes and subtropes listed in the relationship tool were also recorded; the tool is only available on trope pages.

General Results

The final sample included 34 trope pages, nine work pages, five indexes, and one disambiguation page. Of the tropes, 29 were objective, five were subjective, and one was trivia. None of the three forms of the folk ontologies (see also, index, relationship tool) were used on every single page. However, indexes were the most common, with only one page out of the 50 lacking any index, and an average of three indexes per page. See also appeared about 75% of the time, with indexes being much more likely to lack a see also page (4 out of 5) than trope pages (6 out of 35). The relationship tool was only used for about 25% of trope pages. Table 2 summarizes the captured data, showing the mean, median, mode, minimum, and maximum for each aspect recorded.

	See also	Indexes	Creation date	Number of edits	Wicks
mean	2.8	3.48	Mar-10	144.1	446.96
median	2	3	Jun-10	70	78
mode	0	1	Jun-10	18	28
minimum	0	0	May-07	1	3
maximum	11	15	May-12	646	3248

	Inbounds	Redirects	Relationship tool	Page length
mean	358.08	0.58	0.343	67519.2
median	54	0	0	36081.5
mode	21	0	0	No mode
minimum	3	0	0	1355
maximum	4734	3	2	375766

Table 2: TV Tropes data summary

Observations

As with Wikipedia, one of the first things taken into account was whether or not the usage of one form of folk ontology (see also, index, relationship tool) correlated with usage of any of the other forms of folk ontology. Indexes and see alsos are positively correlated: the more indexes a page is on, the more tropes are likely to be mentioned in its see also section ($r=0.602$, $p<0.05$; see Figure 16). This suggests that when people try to make a page easier to find, they pay attention to both indexing and the see also section of related tropes. One might think that the effect is due to the age of the articles—older articles have simply accumulated more indexes and related tropes over time—but this seems unlikely to actually be the cause of correlation, as indices showed almost zero correlation with creation date ($r=0.001$), even though see alsos were significantly correlated with creation date ($r=-0.474$, $p<0.05$; so older articles were more likely to have larger see also sections).

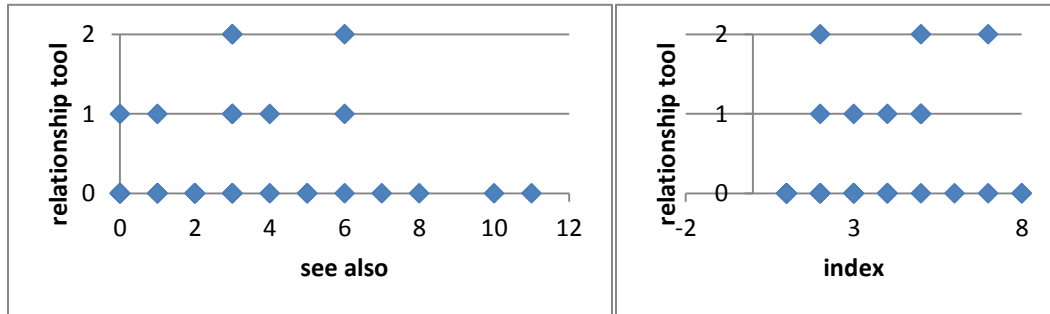
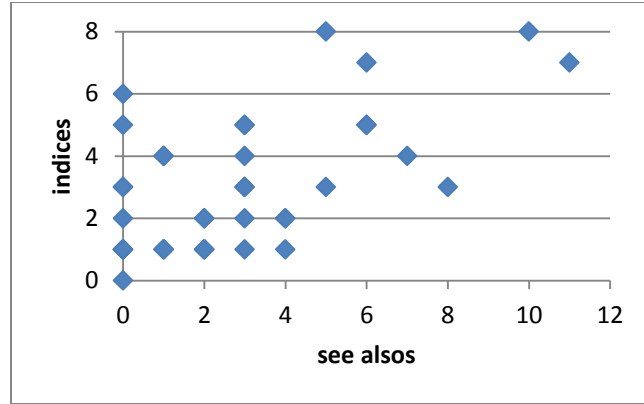


Figure 16: the forms of folk ontologies relative to each other

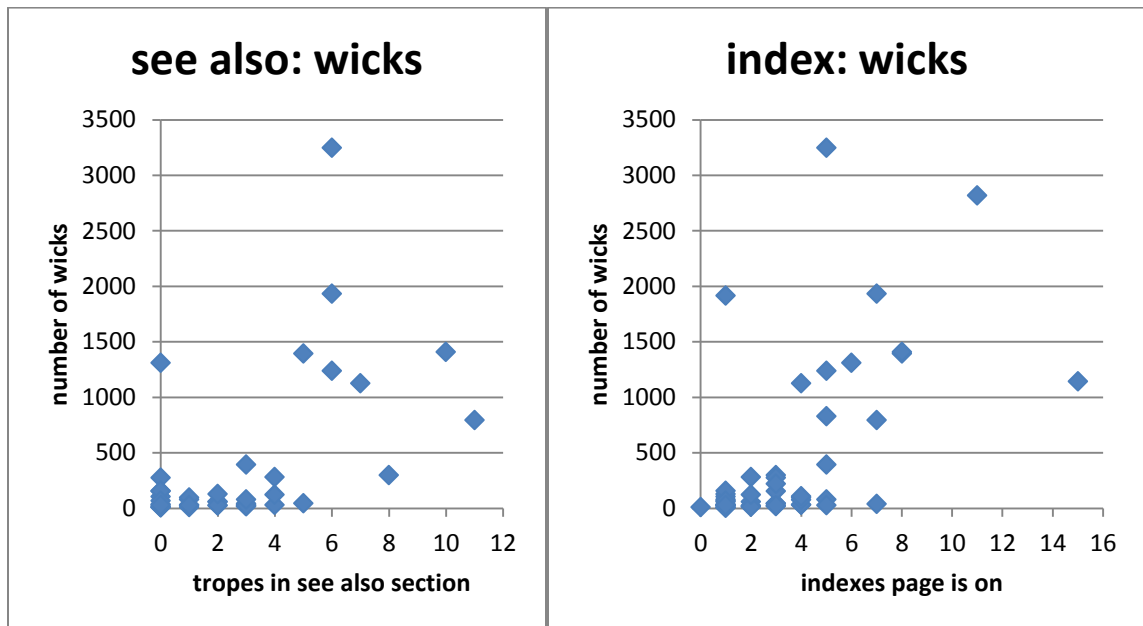


Figure 17: see also sections and indices relative to wicks

Moving on to other parts of the gathered data, there is a positive correlation between wicks and both the use of see also sections and the use of indexes (see also: $r=0.537$, $p<0.05$; indices: $r=0.582$, $p<0.05$; see Figure 17). This suggests that either both those forms of folk ontologies are useful for attracting editor attention—the more aware editors are of the pages, the more likely they are to link to them from other pages—or that when editors pay attention to a page, they focus on its wicks, indexes, and relationships.

Suggesting similar things as the correlation between wicks and the folk ontologies does, the number of edits on a given page is also positively correlated with both see also sections and indexes (see also: $r=0.575$, $p<0.05$; indices: $r=0.456$, $p<0.05$; see Figure 18). For indexes, the correlation holds best for trope pages ($r=0.709$, $p<0.05$) and falls apart when considering just work pages ($r=0.384$, $p>0.05$). This can be seen on the chart, as the two outliers are both work pages. As with wicks, a greater number of edits indicates greater editor attention on the pages; the fact that larger see also sections or more indexes seem to predict more edits once again suggests that either those two forms of folk ontologies are useful for drawing editor attention to the pages or attention to the page causes those forms to be expanded upon.

Following on from this trend, those same two forms of folk ontology are also correlated with longer page length (see also: $r=0.541$, $p<0.05$; indices: $r=0.546$, $p<0.05$; see Figure 19). The one seemingly glaring anomaly in the index chart is actually somewhat deceptive—the work page with 11 indexes had actually grown so large that its trope list (which usually forms the bulk of the page) had to be split off onto subpages, as the wiki software breaks if pages exceed 500,000 characters. When coding the data, only the size of the main page itself, holding just a description of the work and links to the

subpages, was taken into account, which is why it appears so small. As with wicks and page edits, page length is also indicative of editor attention. Editing attention itself is likely indicative of reader attention more generally, as most editors also read the wiki regularly. As we have no way of directly measuring the number of page views, this is likely the closest metric available for estimating general reader attention, which suggests that both see also sections and indexes are fairly strongly correlated with reader attention.

One other possible measure of gauging general reader attention is inbounds, which are the number of hits the page has received via off-site, non-search-engine links. Inbounds mean that people are linking to the wiki from out in the web; in order for a page to be linked to, readers must be aware of its existence. So inbounds are at least slightly reflective of the number of people who have read a given page. All three forms of folk ontology are correlated with the number of inbounds (see also: $r=0.413$, $p<0.05$; indices: $r=0.373$, $p<0.05$; relationship tool: $r=0.410$, $p<0.05$; see Figure 20).

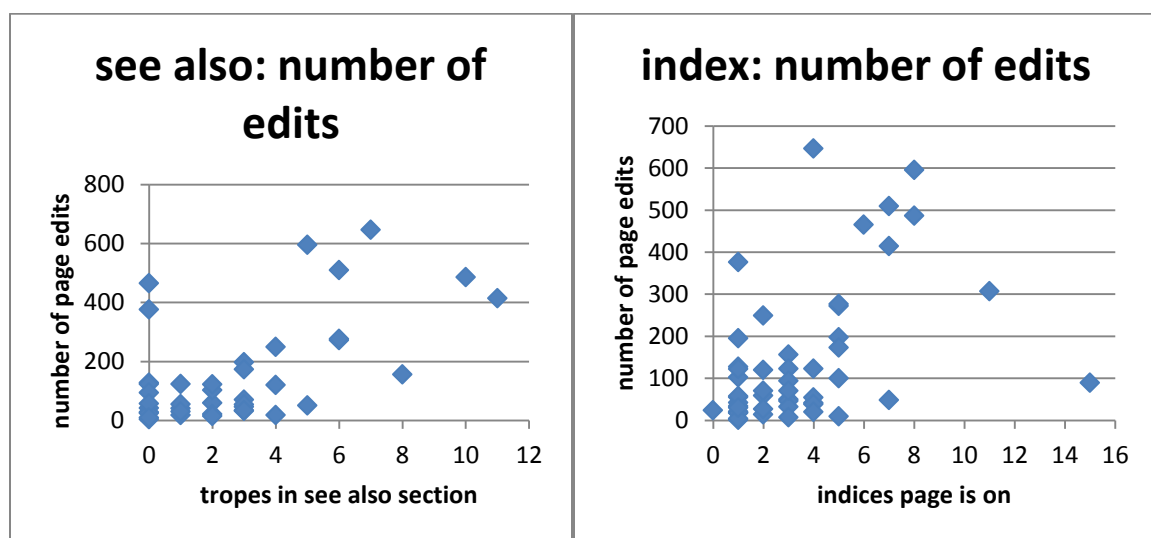


Figure 18: see also sections and indices relative to the number of times a page has been edited

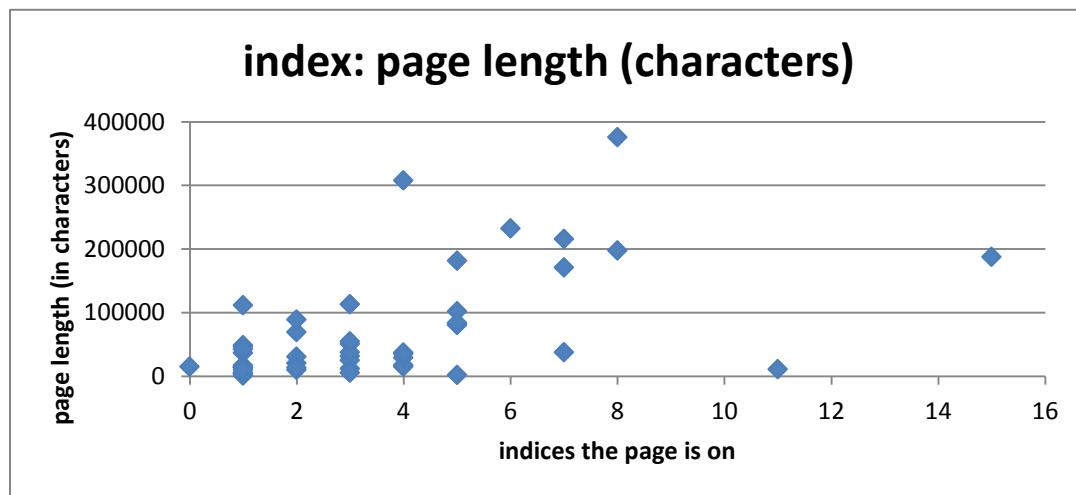
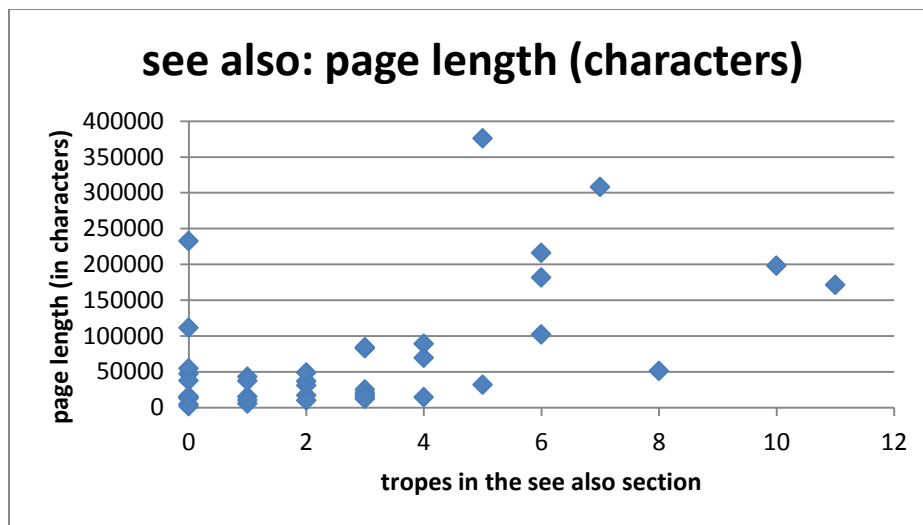


Figure 19: see also sections and indices relative to page length (in characters)

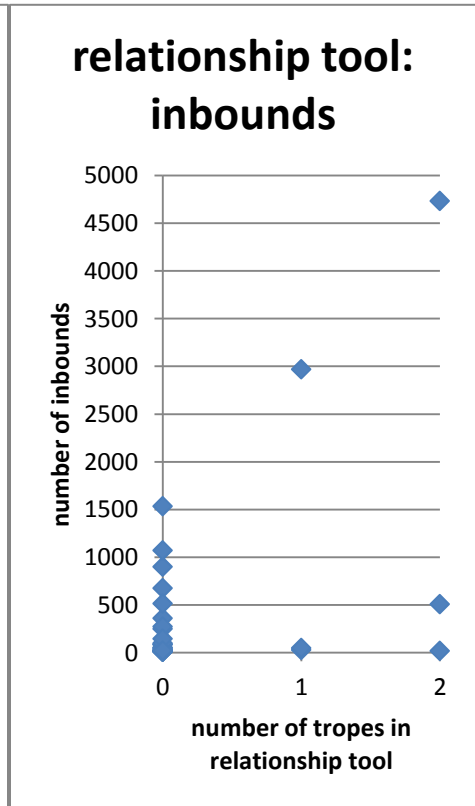
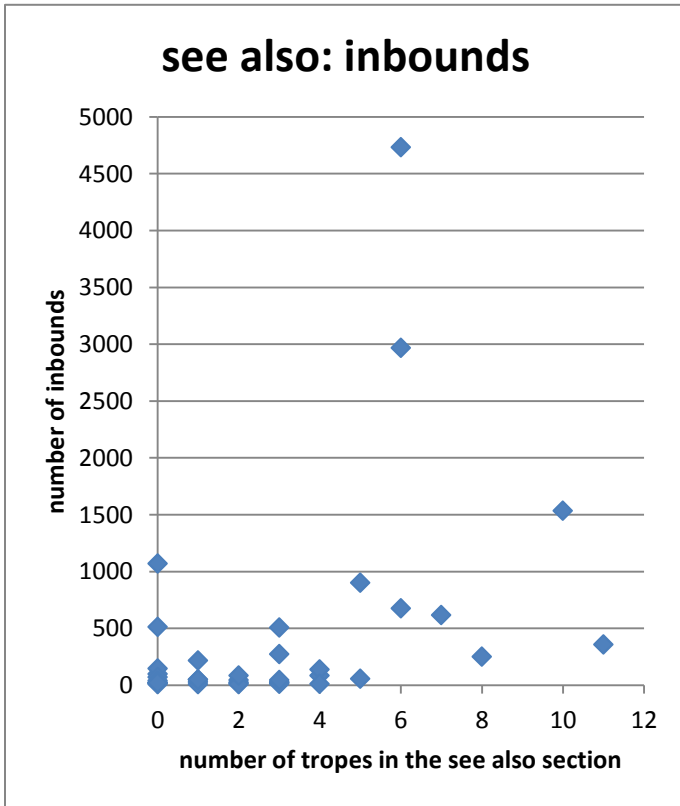
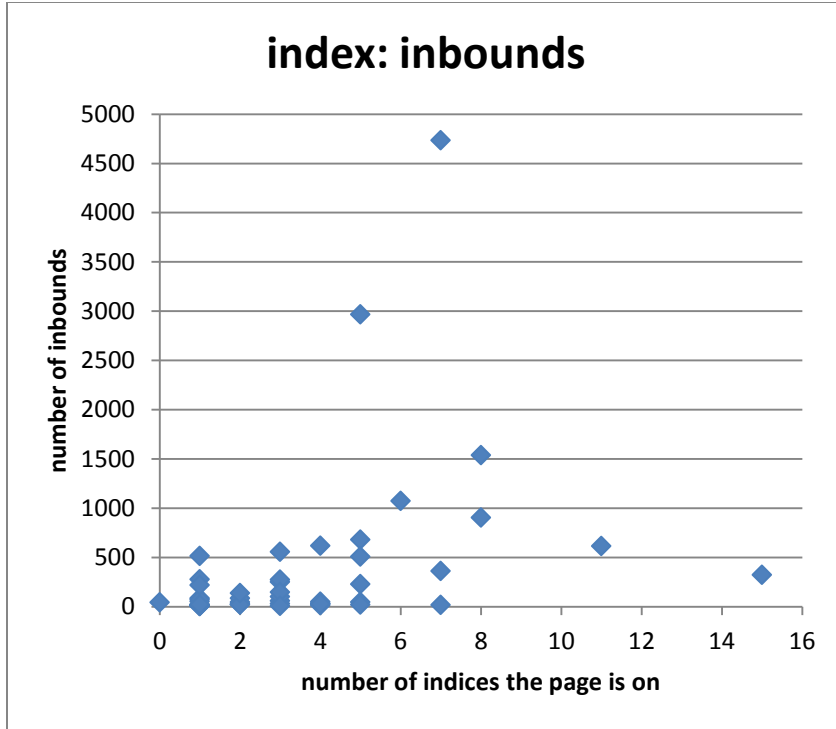


Figure 20: inbounds relative to the three forms of folk ontology

One thing which may be noticed from this section is the paucity of findings related to the relationship tool. This is because the use of the relationship tool is not significantly correlated with any of the measures used aside from inbounds, suggesting that it either has very little impact on helping people find pages on TV Tropes or that editors either do not know about it or do not care about it. In contrast, both see also sections and indexes showed clear correlation with measures of reader and editor attention.

Comparison of Findings Between Wikipedia and TV Tropes

Wikipedia and TV Tropes have some notable differences: Wikipedia's pages are better documented than TV Tropes', making it easier to collect the data for Wikipedia; and TV Tropes has a much more complex organization of its primary pages than Wikipedia, as can be seen from the methodology sections for each wiki. However, despite these differences, the trends in the data from the two wikis show some striking similarities.

Each wiki had three forms of its folk ontologies broadly accessible across pages (Wikipedia's lists don't count for this, as they are centralized in a single location). Of those three forms, one was used on all or almost all of the pages—this was the one which wiki rules or guidelines require to be used for all pages—one was used on a majority of the pages, and one was used on only a minority of pages. Most likely coincidentally, if the percentage of usage for the latter two forms are added together for each wiki, it totals 100% (two-thirds and one-third for Wikipedia, three-quarters and one-quarter for TV Tropes). In both wikis, usage of one of the two most-used forms of folk ontologies was correlated with usage of the other most-used form of folk ontology. It is difficult to tell

whether the use of one type encourages the use of the other, or if they just happen to co-occur (e.g., editors who give particular attention to the page add both).

Another trend common between the wikis was that the most-used form of the folk ontology was correlated with a wide variety of page viewing and editing indicators, while the least-used form had little to no correlations. This could be indicative of a number of things. It may be possible that broader use of a folk ontology is necessary for most users to take notice of it and start using it, and thus folk ontologies which have not (yet) been widely adopted on the wiki don't get much use. Of course, if that's the case, then you run into the traditional circular issue of needing visibility to increase use, but needing use to increase visibility. It could also indicate that users gravitate to the folk ontologies they find most useful, and so when there's a choice between multiple forms of folk ontology, the ones generally deemed least useful (whether because of barriers to access or layout or something else) are rarely used. Another possibility is that low use is indicative of time elements. Based on creation dates, Wikipedia's navboxes appear to have reached a peak of popularity about five years ago, after which use fell drastically. Culturally, Wikipedia's primary users and editors may have chosen to no longer pay attention to navboxes, for whatever reason, causing their usefulness to drop off even if they are not officially deprecated. On the other hand, TV Tropes' relationship tool was introduced about a year and a half ago, in contrast to the far older indexes and see also sections, so the relationship tool may still be struggling in the beginning stages of adoption and have not yet hit its full stride. Thus, the low use of certain forms of folk ontologies could indicate that the form is either past its prime or yet to reach it. Of course, it's impossible to tell which of these speculations is correct with the data at hand.

Another commonality between the wikis is that when one of the more-used forms of folk ontology occurred for a page, multiple instances of that form were generally present. The exception to this was Wikipedia's sidebars, which usually only had one per page; however, Wikipedia's style guidelines recommend using no more than one sidebar per page, so that undoubtedly affects the frequency of sidebars per page. Categories, indexes, and see also sections averaged two to three of each per page. This highlights the already observed similarities between them.

Thus, despite differences in community and organization, the use of folk ontologies has much in common between the two wikis. This suggests that these findings are likely at least partially generalizable to other instances of folk ontologies.

Conclusion

Summary

This paper took an in-depth look at folk ontologies via two case studies: Wikipedia and TV Tropes. We saw that Wikipedia's primary folk ontology is its category system, with navigational templates and lists forming supplemental folk ontologies. TV Tropes consists of three folk ontologies that somewhat bleed into each other in practice: tropes, works, and indexes. For both wikis, the folk ontologies are maintained by the editors, with major changes being handled according to consensus. Considering the quality of the folk ontologies in light of general ontology evaluation principles, we saw that folk ontologies sacrifice consistency and concision to gain flexibility. Looking at the patterns of use of the manifestations of the folk ontologies on the wikis, it became clear that the most used folk ontologies seem to have a positive impact on the pages being

findable, suggesting that despite the quality concerns, folk ontologies serve their purpose as navigational aids.

Areas for Further Study

There are multiple avenues of further study possible. Broadening the scope to look at more than two wikis would allow results to be more broadly generalized, giving a better feel for how folk ontologies usually are organized and maintained. It would be a good idea particularly to look at some smaller wikis, as those are far more common than large and extremely large wikis like TV Tropes and Wikipedia, respectively. Studies could also look in more depth at one or both of TV Tropes or Wikipedia; a quantitative analysis with a much larger sample size would allow conclusions, rather than assumptions and suppositions, to be drawn about the use and impact of their folk ontologies. Additionally, more work could be performed on articulating a framework for evaluating folk ontologies, and a more detailed analysis could be performed, which provides more depth than the flying overview in this paper.

Final Thoughts

Beyond that, we can consider what the folk ontologies reveal about how people choose to collectively organize information when not guided by information professionals. The fact that, on both Wikipedia and TV Tropes, pages are generally assigned to multiple categories or indexes (respectively) indicates that people tend to pull two to three aspects out as the most important for any given topic, reflecting and reinforcing the traditional cataloging advice that two or three subject headings per record is ideal. Additionally, the categorization or indexing often includes non-hierarchical associations (related concepts or about-a relationships) as well as hierarchical is-a

relationships. The fact that people building folk ontologies find the non-hierarchical relationships important to indicate suggests that using ontologies with multiple types of relationships defined, to provide full richness of relationship description, may best reflect how the average person thinks about information structures.

The mere existence of folk ontologies shows that people in aggregate are capable of producing far more complex information organization than folksonomies. Folk ontologies also highlight the fact that humans, when faced with a wealth of information, naturally sort it to make it easier to understand and access. And in this day and age, everyone is an organizer of information.

Notes

- [1] It may be possible to extract the structure and encode it in a machine-readable format, but doing so is a complex process which is separate from the folk ontology itself. DBPedia (<http://dbpedia.org/About>), extracted from Wikipedia, is probably the best known attempt at this.
- [2] This numerical coincidence is most likely mere coincidence, rather than an indication that wikis usually have three folk ontologies. A more general study across a much greater number of wikis would be needed to verify that in either direction.
- [3] “Categorization.” *Wikipedia*. 6 March 2013.
<<http://en.wikipedia.org/wiki/Wikipedia:Categorization>>.
- [4] “Categories, lists, and navigation templates.” *Wikipedia*. 9 March 2013.
<[http://en.wikipedia.org/wiki/Wikipedia:Categories, lists, and navigation templates](http://en.wikipedia.org/wiki/Wikipedia:Categories,_lists,_and_navigation_templates)>.
- [5] “Fundamental categories.” *Wikipedia*. 4 February 2013.
<http://en.wikipedia.org/wiki/Category:Fundamental_categories>.
- [6] “Main topic classifications.” *Wikipedia*. 30 December 2012.
<http://en.wikipedia.org/wiki/Category:Main_topic_classifications>.
- [7] “Overcategorization.” *Wikipedia*. 11 March 2013.
<<http://en.wikipedia.org/wiki/Wikipedia:Overcategorization>>.
- [8] “Page Type Counts.” *TV Tropes*. 19 March 2013.
<http://tvtropes.org/pmwiki/page_type_counts.php>.

[9] “Work Pages Are a Free Launch.” *TV Tropes*. 27 December 2012.

<<http://tvtropes.org/pmwiki/pmwiki.php/Administrivia/WorksPagesAreAFreeLaunch>>.

[10] “Missing Supertrope”. *TV Tropes*. 12 March 2013.

<<http://tvtropes.org/pmwiki/pmwiki.php/Main/MissingSupertrope>>.

[11] “Index Index.” *TV Tropes*. 13 March 2013.

<http://tvtropes.org/pmwiki/index_report.php>.

[12] “Audience Reactions.” *TV Tropes*. 4 March 2013.

<<http://tvtropes.org/pmwiki/pmwiki.php/Main/AudienceReactions>>.

[13] “YMMV.” *TV Tropes*. 1 March 2013.

<<http://tvtropes.org/pmwiki/pmwiki.php/Main/YMMV>>.

Bibliography

- Auer, Sören, et al. "DBpedia: A Nucleus for a Web of Open Data." *The Semantic Web*. Eds. Karl Aberer, et al. Berlin: Springer, 2007. 722-35.
- Di Ioria, Angelo, Alberto Musettia, Silvio Peronia, and Fabio Vitali. "Ontology-driven Generation of Wiki Content and Interfaces." *New Review of Hypermedia and Multimedia* 16.1-2 (2010): 9-31.
- Gómez-Pérez, Asunción. "Evaluation of Ontologies." *International Journal of Intelligent Systems* 16 (2001): 391-409. Wiley DOI: 10.1002/1098-111X(200103)16:3<391::AID-INT1014>3.0.CO;2-2.
- Kousetti, Chrysovalanto, David E. Millard, and Yvonne Howard. "A Study of Ontology Convergence in a Semantic Wiki." In *Proceedings of the 4th International Symposium on Wikis (WikiSym '08)*. New York: ACM, 2008.
- Kuhn, Tobias. "AceWiki: Collaborative Ontology Management in Controlled Natural Language." In *Proceedings of the 3rd Semantic Wiki Workshop, CEUR Workshop Proceedings, 2008*. *arXiv.org*. DOI: arXiv:0807.4623.
- Liu, Jiahui, and Daniel M. Gruen. "Between Ontology and Folksonomy: A Study of Collaborative and Implicit Ontology Evolution." In *Proceedings of the 13th international conference on Intelligent user interfaces (IUI '08)*. New York: ACM, 2008. 361-64.
- Maleewong, Krissada, Chutiporn Anutariya, and Vilas Wuwongse. "A semantic argumentation approach to collaborative ontology engineering." In *Proceedings*

of the 11th International Conference on Information Integration and Web-based Applications & Services (iiWAS '09). New York: ACM, 2009. 56-63.

Medelyan, Olena, David Milne, Catherine Legg, and Ian H. Witten. "Mining meaning from Wikipedia." *International Journal of Human-Computer Studies* 67 (2009): 716-54.

Navigli, Roberto, Paola Velardi, Alessandro Cucchiarelli, and Franchesca Neri.

"Quantitative and Qualitative Evaluation of the OntoLearn Ontology Learning System." In *Proceedings of the 20th international conference on Computational Linguistics (COLING '04)* (Article No. 1043). ACM DOI: 10.3115/1220355.1220505.

Noy, Natalya F., Abhita Chugh, William Liu, and Mark A. Musen. "A Framework for Ontology Evolution in Collaborative Environments." *The Semantic Web - ISWC 2006*. Eds. Isabel Cruz, et al. Berlin: Springer, 2006. 544-58.

Palma, Raúl, Oscar Corcho, Asunción Gómez-Pérez, and Peter Haase. "Web Semantics: Science, Services and Agents on the World Wide Web." *Semantic Web Dynamics* 9 (2011): 299-314.

Pak, Jinie, and Lina Zhou. "A Framework for Ontology Evaluation." *Exploring the Grand Challenges for Next Generation E-Business*. Eds. Raj Sharman, H. Raghav Rao, and T. S. Raghu. Berlin: Springer, 2010. 10-18.

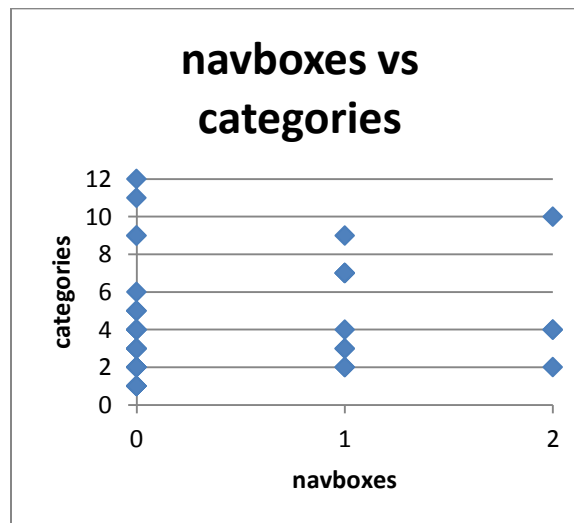
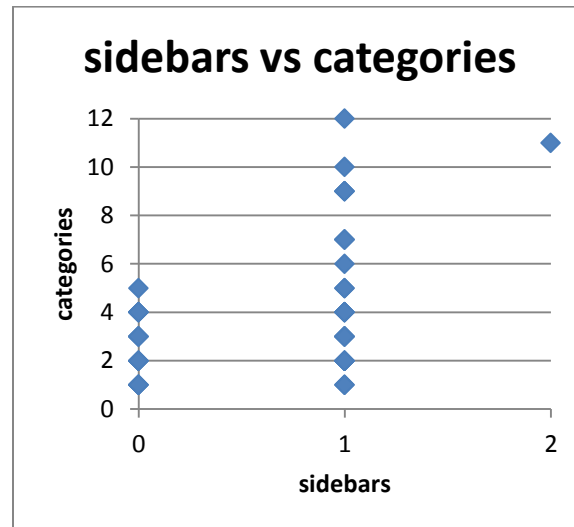
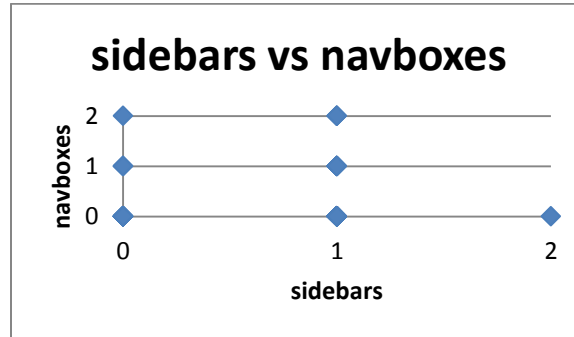
Ponzetto, Simone Paolo, Michael Strube. "Knowledge Derived From Wikipedia For Computing Semantic Relatedness." *Journal of Artificial Intelligence Research* 30 (2007): 181-212.

- Sabou, Marta, and Miriam Fernandez. "Ontology (Network) Evaluation." *Ontology Engineering in a Networked World*. Eds. Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, Enrico Motta, and Aldo Gangemi. Heidelberg: Springer, 2012. 193-212.
- Solskinnsbakk, Geir, et al. "Quality of Hierarchies in Ontologies and Folksonomies." *Data & Knowledge Engineering* 74 (2012): 13-25.
- Suchanek, Fabian M., Gjergji Kasneci, and Gerhard Weikum. "YAGO: A Large Ontology from Wikipedia and WordNet." *Web Semantics: Science, Services and Agents on the World Wide Web* 6 (2008): 203-17.
- TV Tropes*. <<http://www.tvtropes.org>>.
- Vrandečić, Denny. "Ontology Evaluation." *Handbook on Ontologies*. 2nd ed. Eds. Steffen Staab and Rudi Studer. Dordrecht: Springer, 2009. 293-313.
- Wikipedia*. <<http://www.wikipedia.org>>.
- Wu, Fei, and Daniel S. Weld. "Automatically Refining the Wikipedia Infobox Ontology." WWW Conference 2008. Accessed 21 March 2013
<<http://wwwconference.org/www2008/papers/pdf/p635-wu.pdf>>.

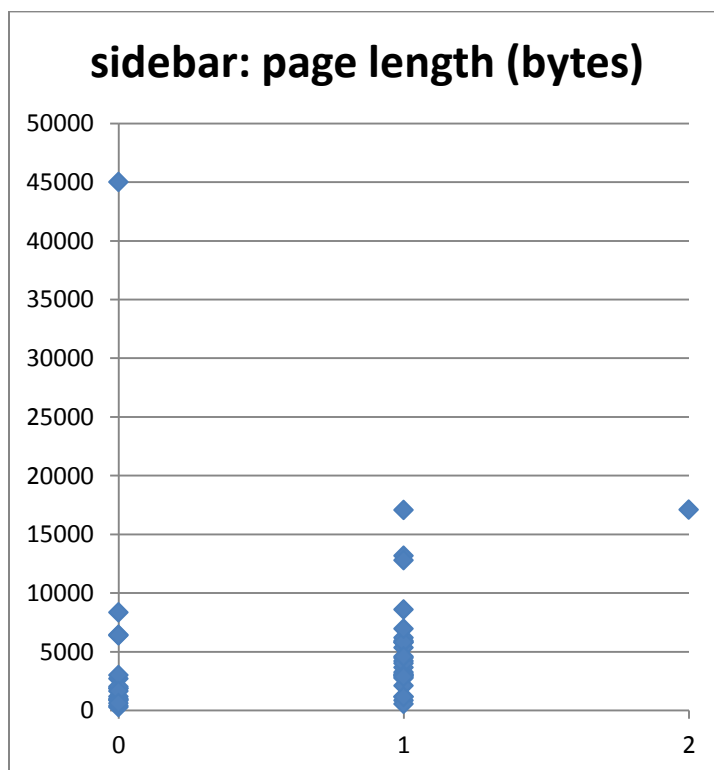
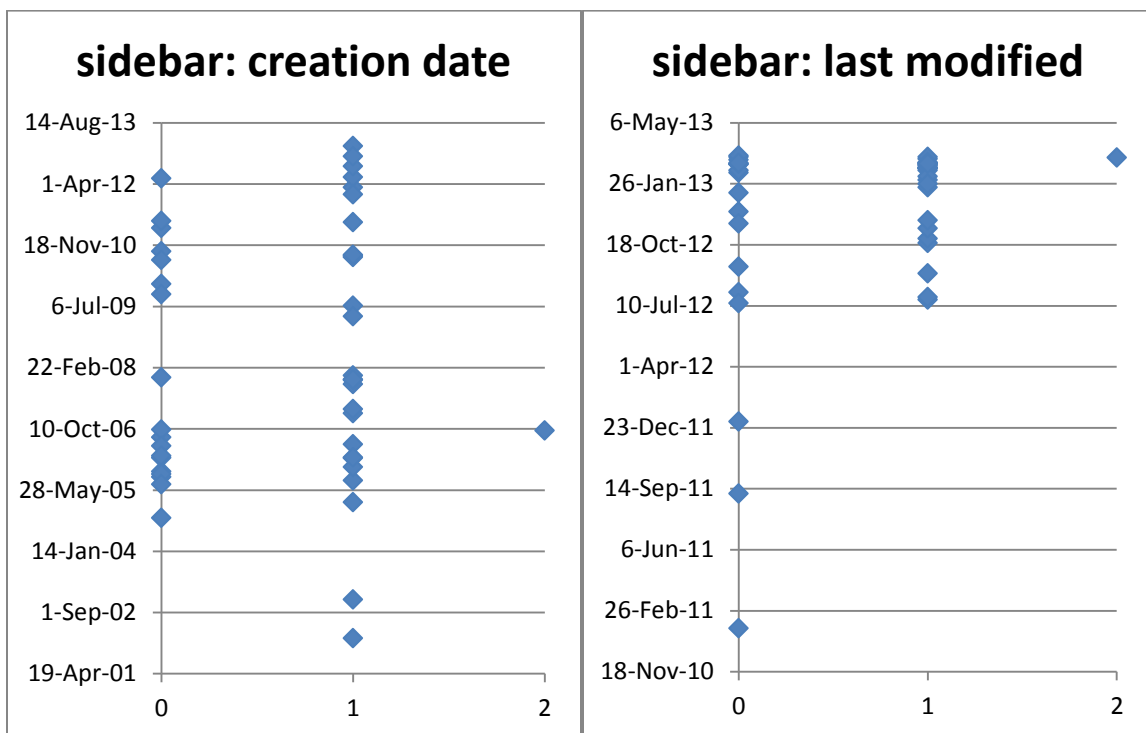
Appendix 1: Graphs from the Quantitative Analysis

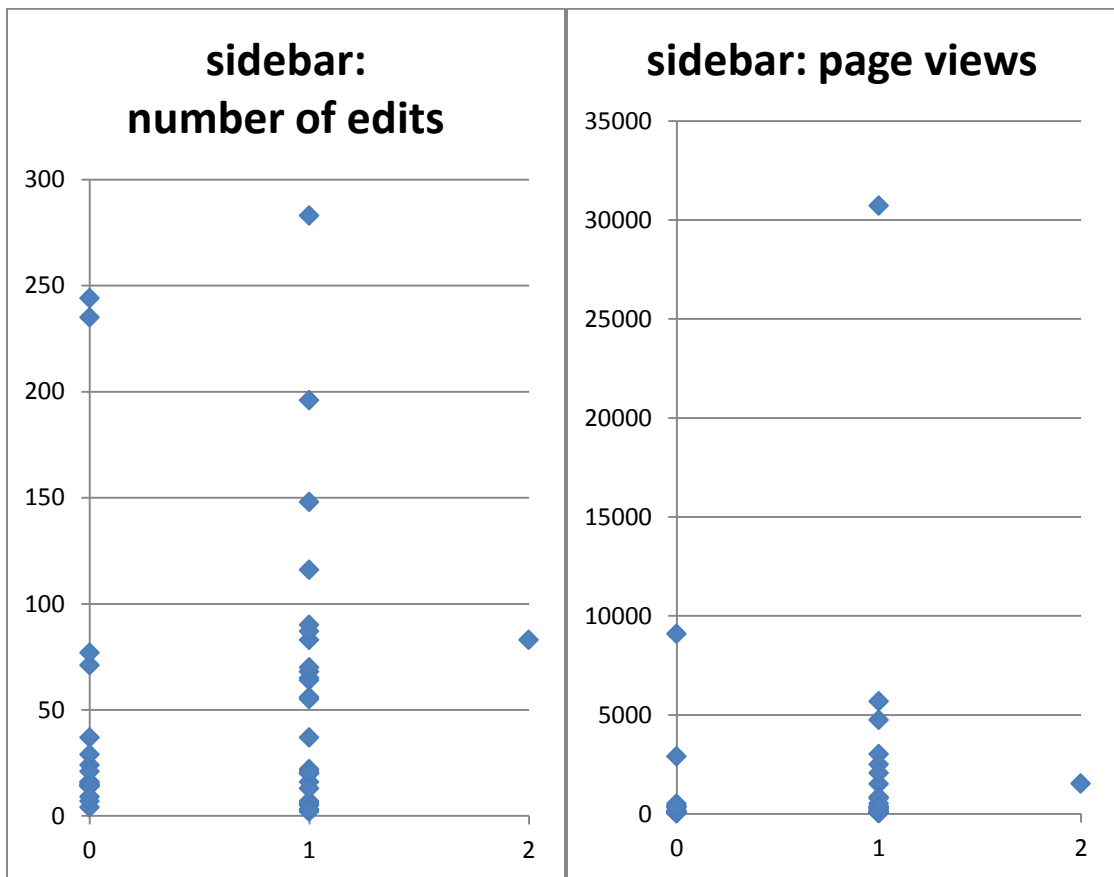
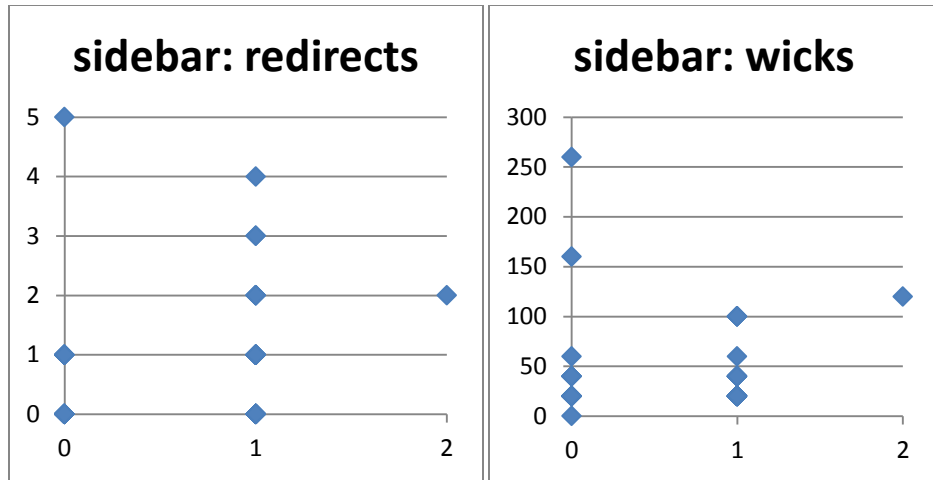
Wikipedia

Usage of the Three Folk Ontologies Enumerations Relative to Each Other

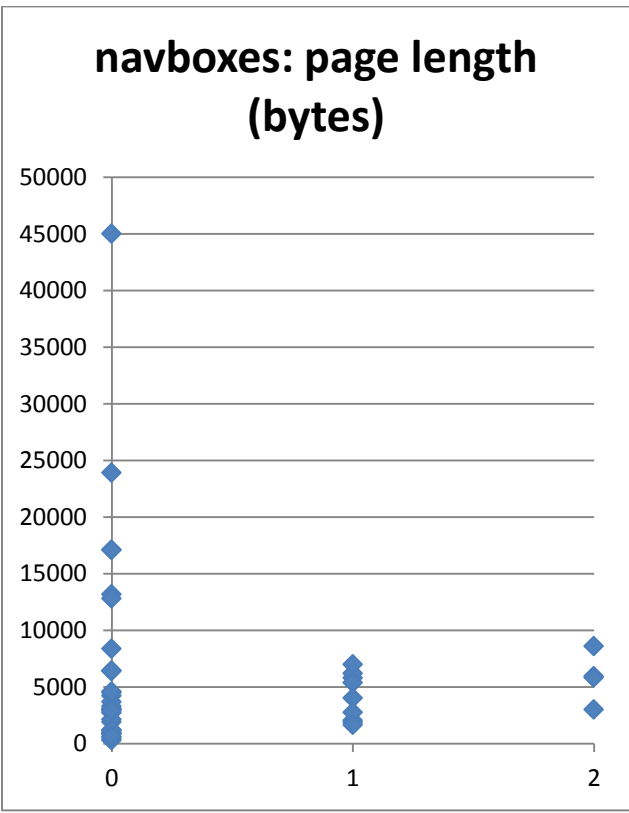
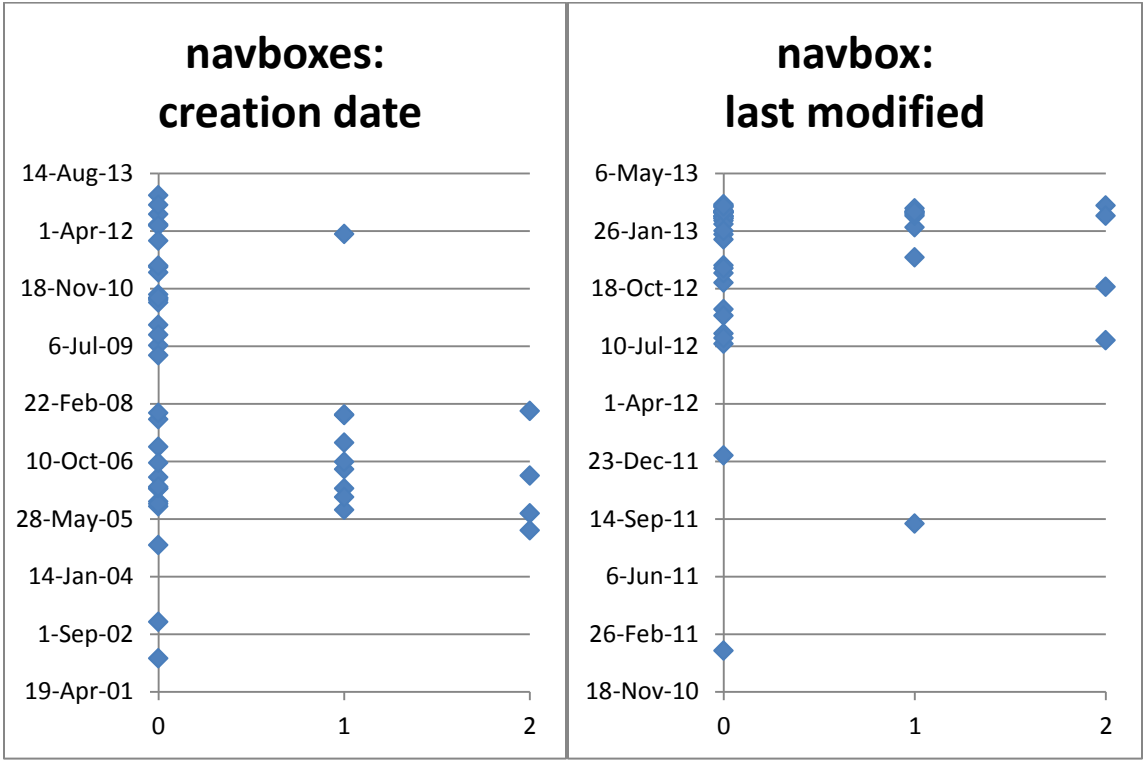


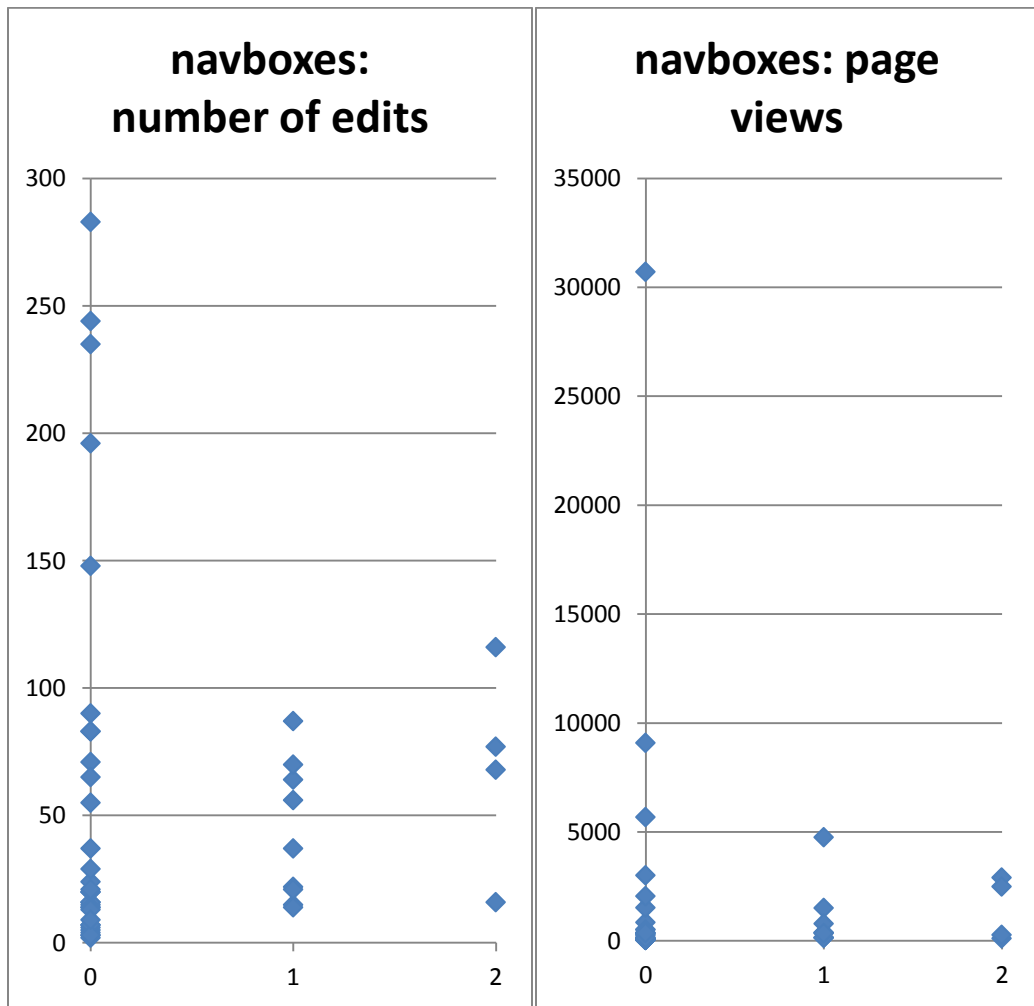
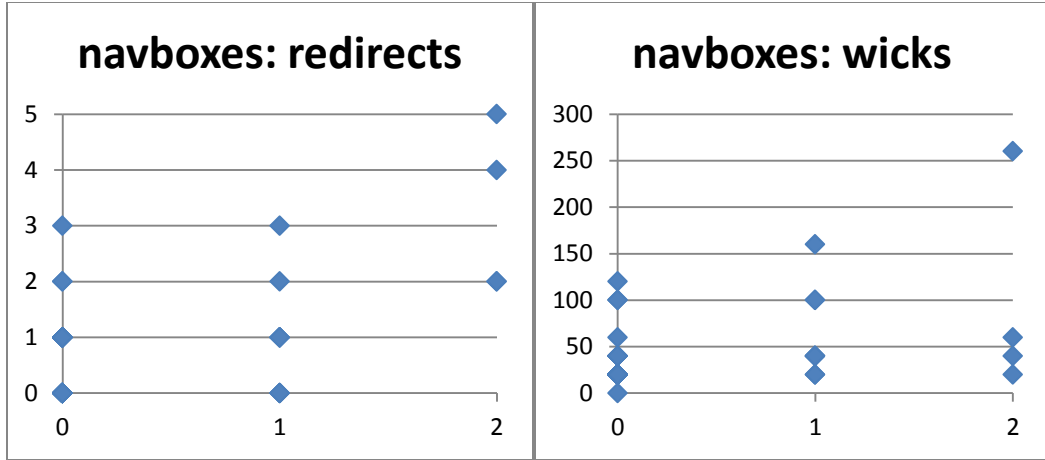
Sidebars' Correspondences



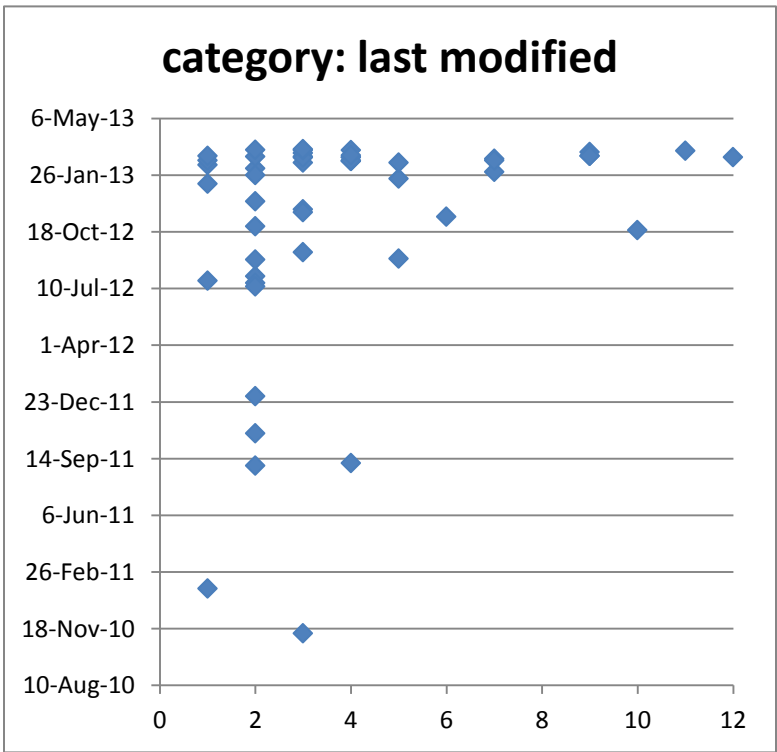
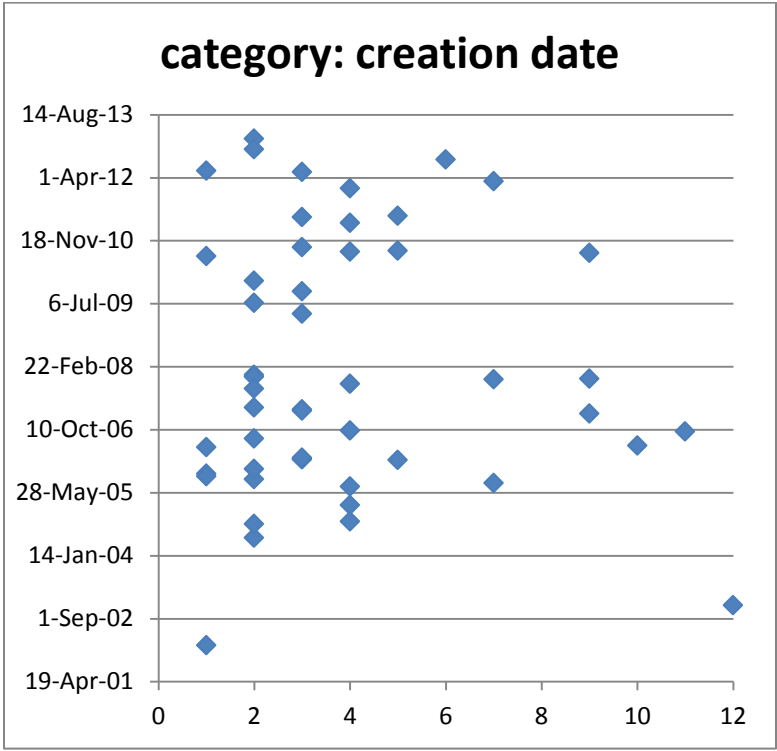


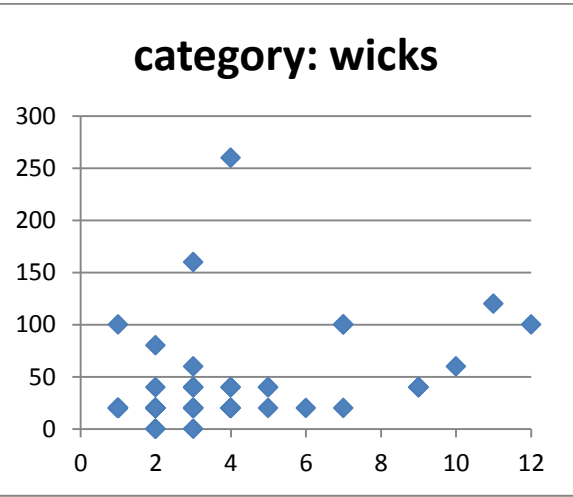
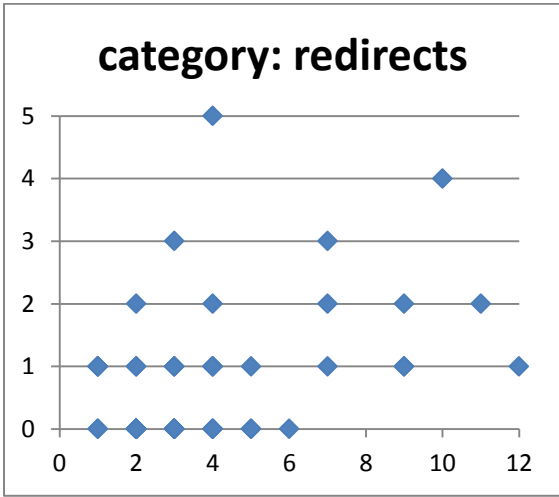
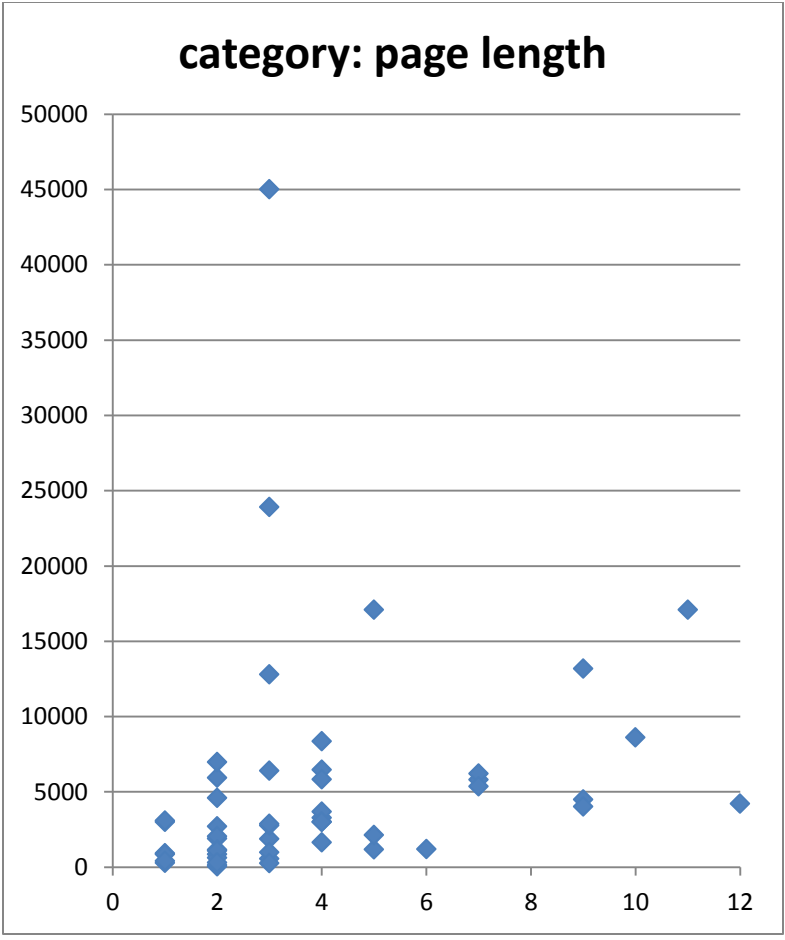
Navboxes' Correspondences

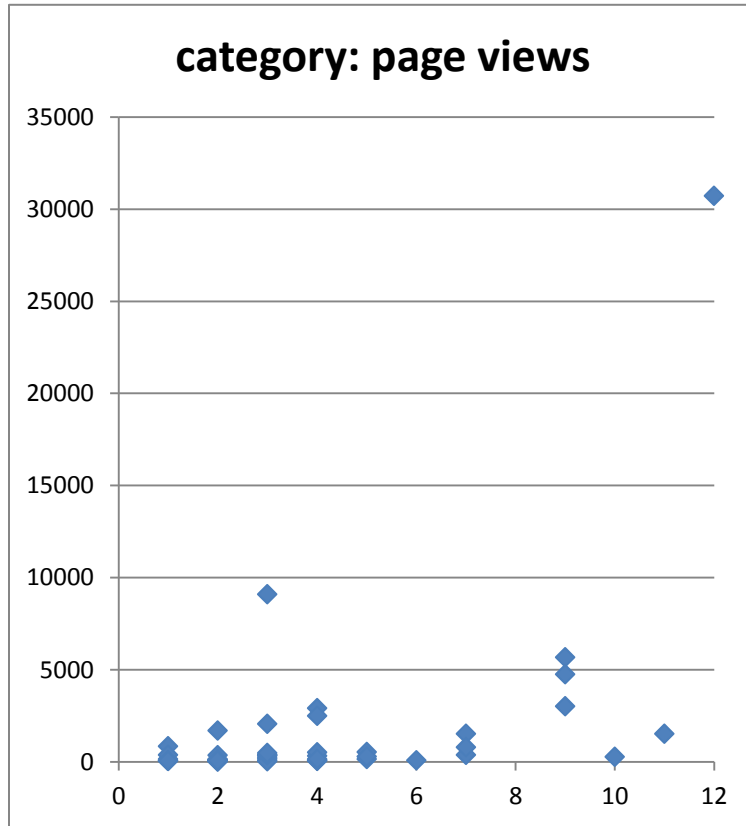
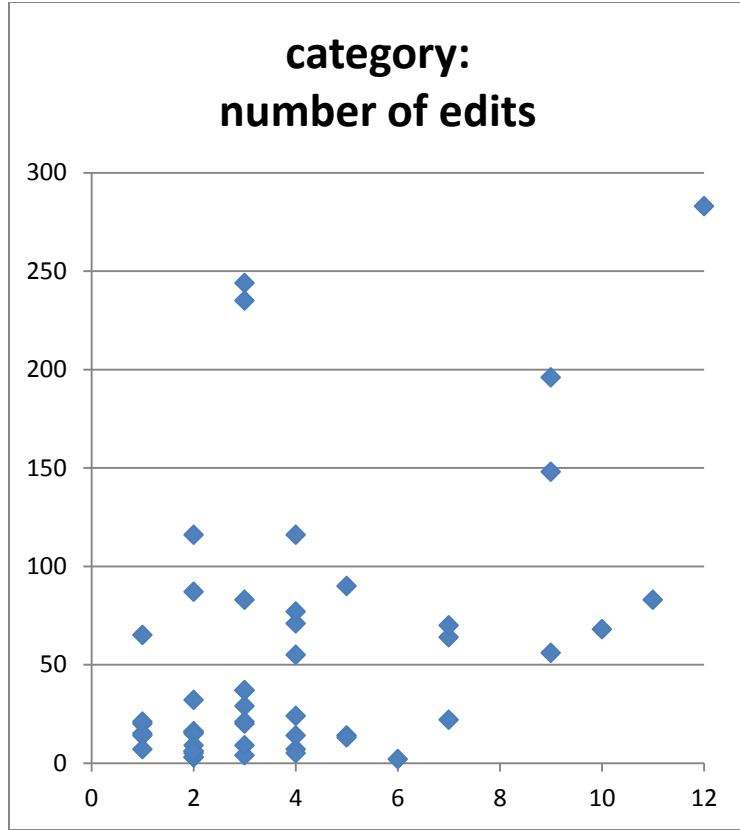




Categories' Correspondences

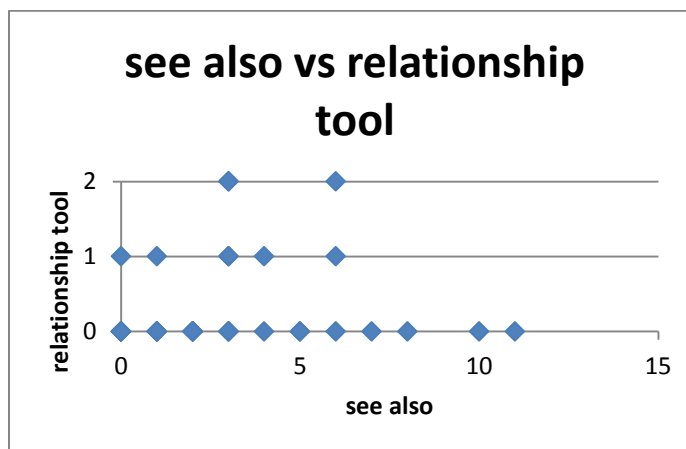
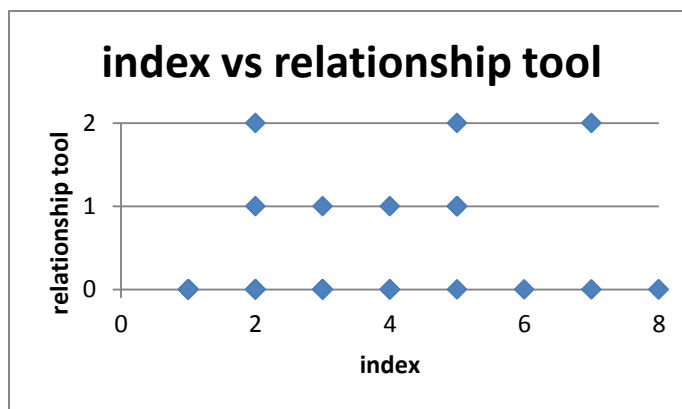
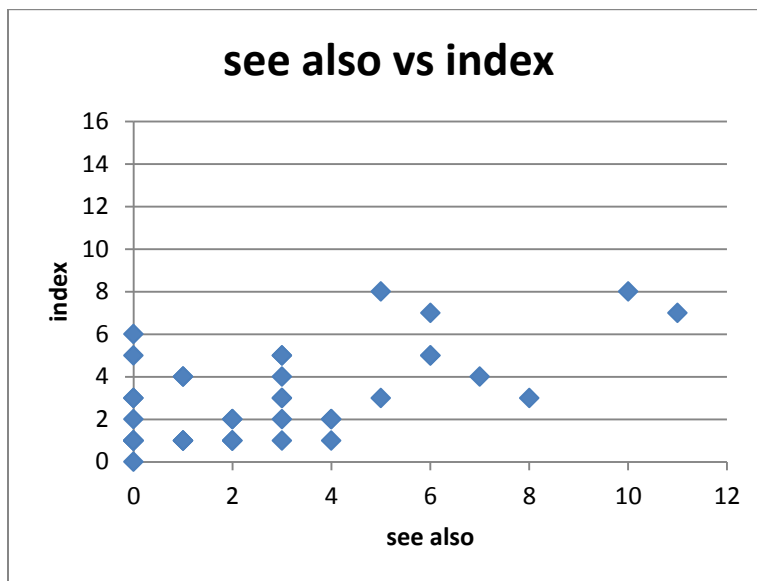




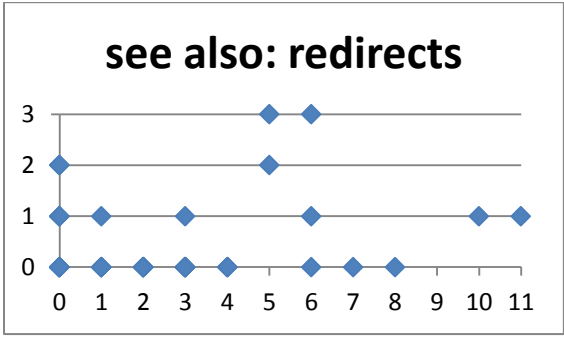
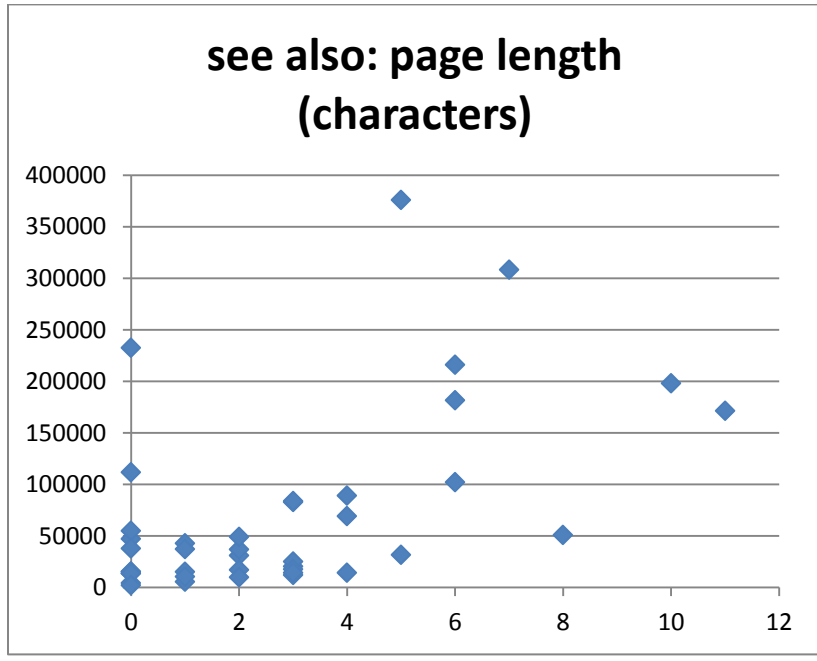
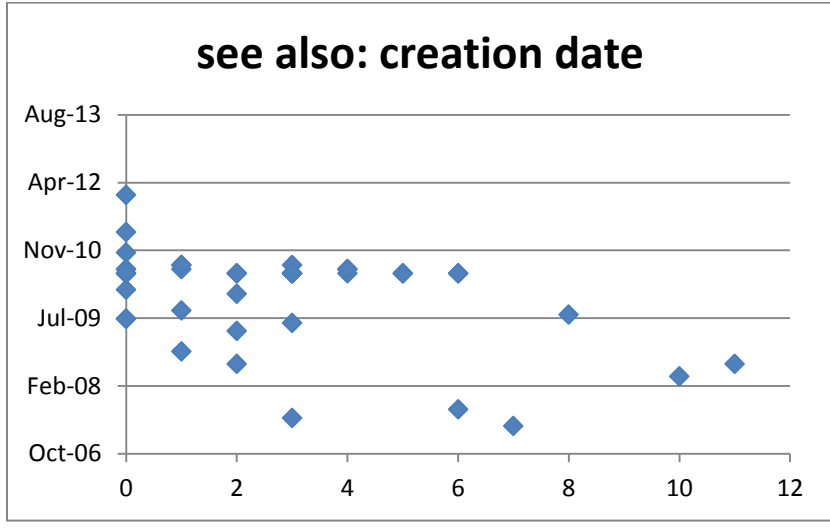


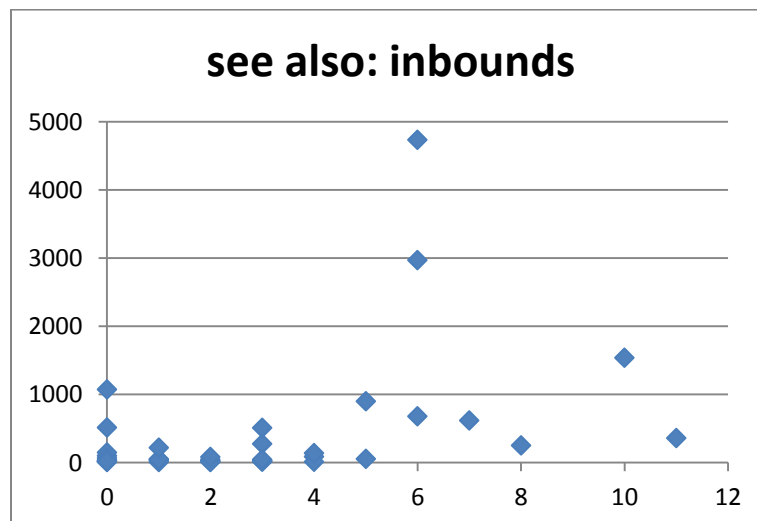
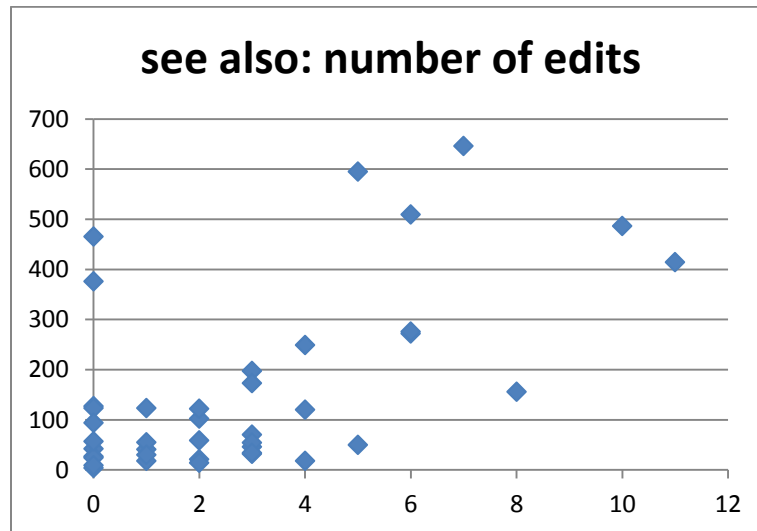
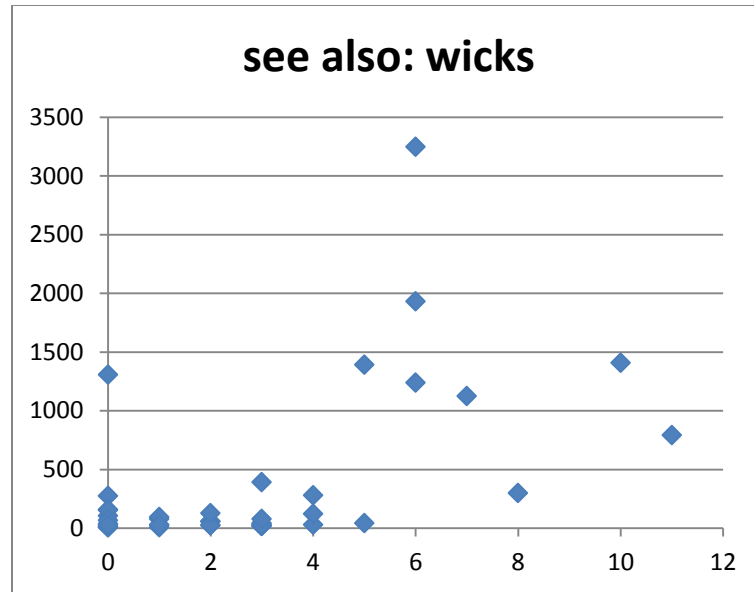
TV Tropes

Usage of the Three Folk Ontology Enumerations Relative to Each Other

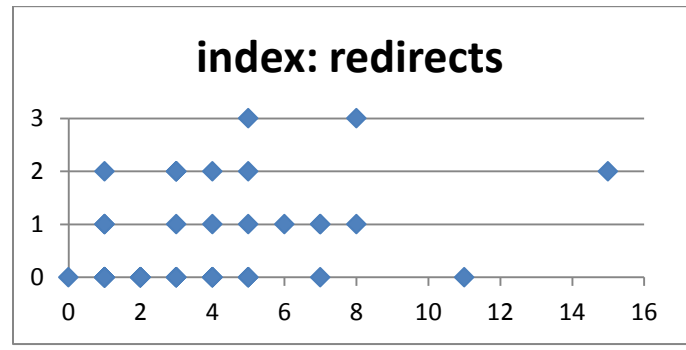
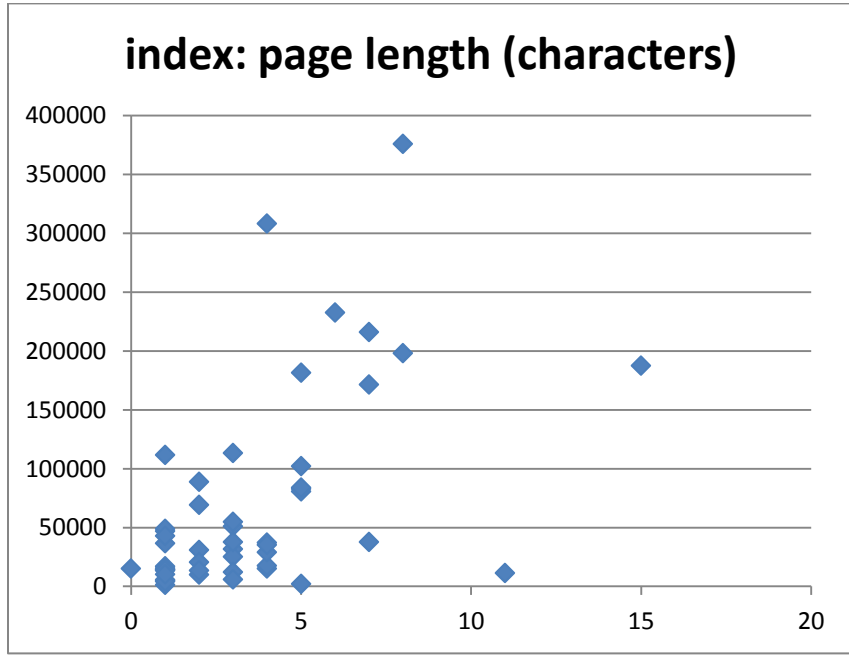
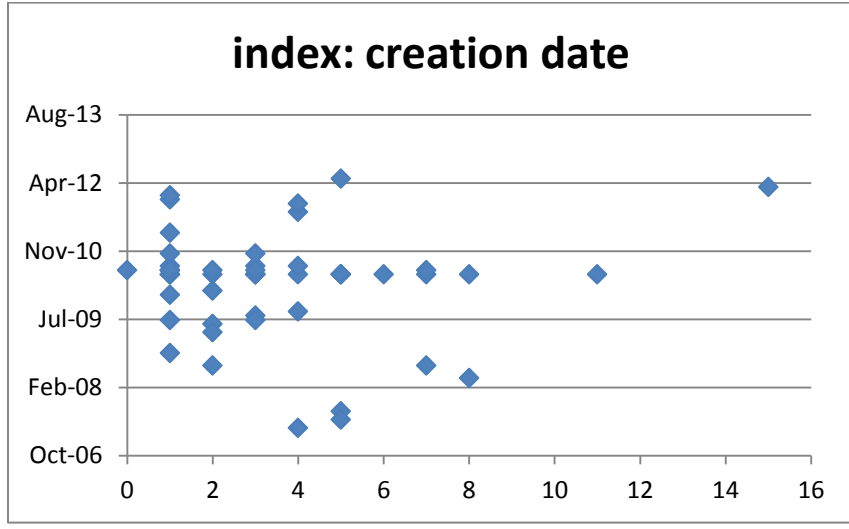


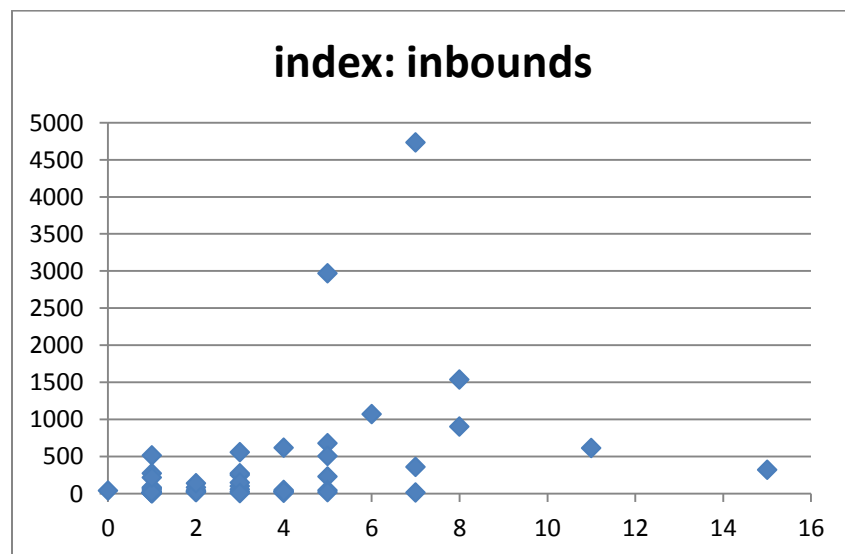
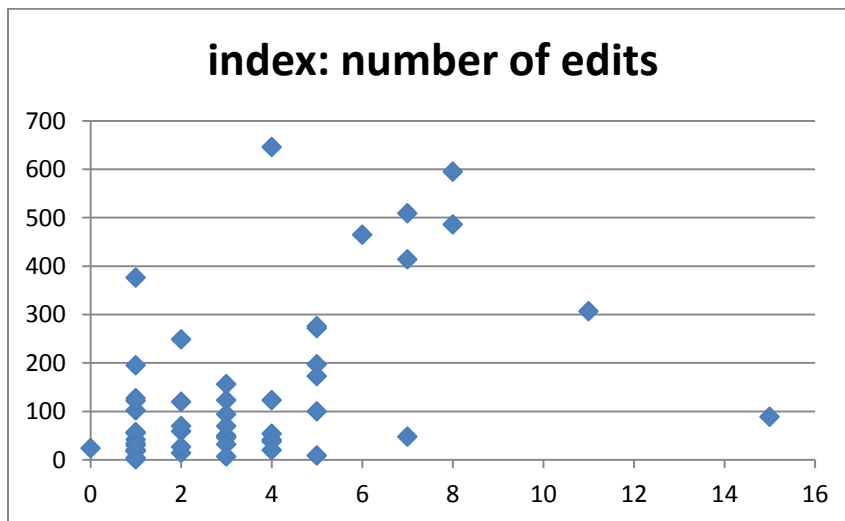
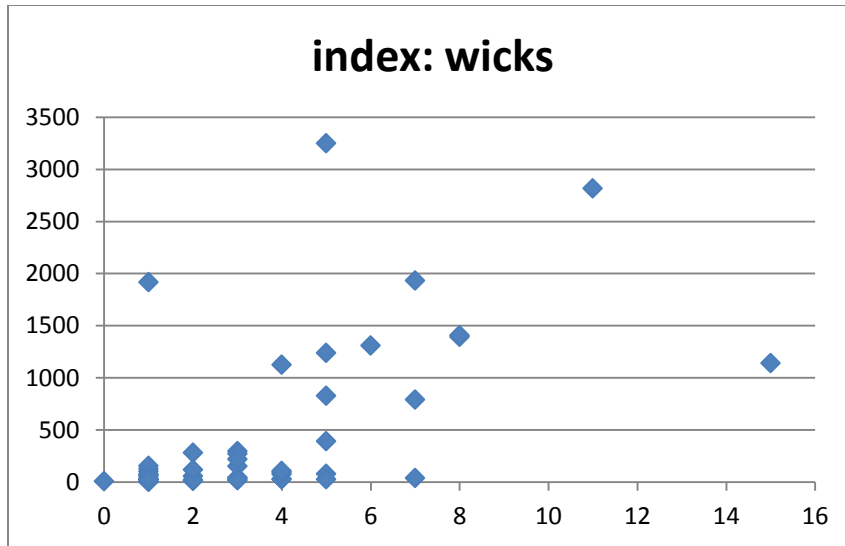
See Also Areas' Correspondences





Indices' Correspondences





Relationship Tool's Correspondences

