Ontology is over-rated: categories, links and tags. By Clay Shirky

Introduction

Taxonomy, Ontology and Folksonomy; my first objective is to describe these three terms, which I have constantly found during my research of the paper by Clay Shirky – Ontology is over-rated: categories, links and tags.

Taxonomy comes from the Greek meanings, Taxis; classification and nomy; management. (Emanuele Quintarelli, 2005).

Ontology can be divided this way, Onto: being or existence and logy: science, doctrine or theory (Webster’s, 1999). A collection of names for a particular concept, like our music instruments assignment and relation types, organised in a type, sub-type and sub-sub type as so on.

Folksonomy is the merging of two words; Folk meaning people, and Taxonomy. Another way of looking at the terms is if taxonomy is like the whole body of a classification system, then ontology is assigning body names to the skeleton, and then making those chosen items fit onto the skeleton, and folksonomy is a bit like anyone naming the skeleton's body parts, according to their own way of describing something, and then the most popular names are used.

Cataloguing and classification

Classification has been happening since man could think; from food they could eat and the food that could eat them, to the weapons they could use to kill the animals; close weapons like clubs or spears or long distance weapons, such as boomerangs or javelins.

Traditional cataloguing and classification has been structured using a hierarchical taxonomy, with specialist ontologists, studying a particular domain, and assigning names, rules and the relationships to those names. These systems are at present found in places such as libraries and museums. The reason for their existence is so that large amounts of material can sorted, stored and retrieved easily by anyone wishing to access it.

Rules were set up for the purposes of being able to sort through the conglomeration of material that is available; in essence “is a response to physical constraints on storage, and to people’s inability to keep the location of more than a few hundred things in their mind at once.” (Shirky, 2005) It becomes evident throughout Shirky’s paper that he is not enthusiastic about the traditional hierarchical scheme. He goes on to say “any concept is supposed to fit in one category and in no other category so that a book may find a place on the shelf, but when a book is not about one main thing it has to be “declared to be about one thing more than others, regardless of its actual contents.” (Shirky, 2005). I think this is
a rather extreme point of view; most books are generally on a topic, have a theme that can be categorised for classification purposes, and within context of the actual library itself.

All cataloguing schemes when created where set, and therefore does not allow for variances easily, changes in the world are not easily adaptive to the traditional classification schemes. Two examples of traditional cataloguing is the Dewey Decimal System developed in 1870 by Melvil Dewey, (http://www.oclc.org/dewey) and the Library of Congress Cataloguing system in 1901 (http://www.loc.gov/cds/index.html) this does bring about some problems, bias is present and the inability to foresee the future. When Dewey developed the Dewey Decimal System, Christianity was the major religion according to his view, and the classification scheme reflects this. From 200 to 290 is devoted to Christianity and from 290 to 299, clumped in this small area, are all other religions, obviously this does not reflect today’s world where all religions are represented equally. Fortune telling or predicting the future is impossible for formal systems, for example, Dewey did not foresee the invention of the computer or the collapse of East Germany, so what to do with these? Squeeze in numbers for computers somewhere, and unify the numbers for East Germany and Germany?

Shirky fails to point out that the Dewey Decimal System is being constantly revised and scrutinised, currently into its 22nd edition, by a specialist group of people, whose purpose is to constantly monitor and revise the current system. “OCLC Online Computer Library Center is a non-profit, membership, computer library service and research organization dedicated to the public purposes of furthering access to the world's information and reducing information costs.” (oclc.org, 2005).

Links

Shirky also only briefly mentions the see and see also references that are available to help direct people in the use of classification systems. This is the reason to establish a controlled thesaurus and vocabulary to best guide the user to where the item may be located, and to the cataloguer as to where to store the item, the earliest form of links.

Yahoo was developed about 10 years ago to try to deal with the huge information that was available of the web, the Yahoo directory, Shirky argues is also a library cataloguing scheme, a file system “with symbolic links, aliases, shortcuts, whatever you want to call them.”, Yahoo hired specialists who devised the directory, with headings, sub-headings, and so on, in fact, in the virtual world “Adding back the shelf” (Shirky, 2005) but with many links to other areas.
The Web Chaos
While cataloguing and classification works for physical organisation of material; with the emergence of
the web, and the huge amount of information available, the naïve users, and with no global authority
governing the web a different solution is ultimately needed.

On the Web, the direction is scalability, flexibility, fluidity and simplicity to satisfy the
demanding needs of millions of people with different cultural and social backgrounds all over the
world. Under these circumstances, traditional precise classification scheme become expensive (to
create and maintain) and probably lose the capability to match the user’s way of thinking and
organizing the world. (Quinterelli, 2005)

Tags
The URL is a unique identifier, it give us the capability to direct people to the contents of a web page,
once you have this anyone can label (tag) those pointers, the URL’s.

Tagging or the use of metadata has, primarily in the form of catalogue records, been the domain
of dedicated professionals working with complex; detailed rule sets and vocabularies (Adam Mathes,
2004)

As the World Wide Web became more complex users started to tag the own URL’s without
organisation, structure, or controlled vocabulary. The tags are usually up-to-date and in natural
language that the users themselves can understand as they themselves see the world.

This new concept of tagging then moved on to the aggregation of tagging, which has lead to the
introduction of folksonomy. “Folksonomies are not simply visitors tagging something for personal use:
they also are an aggregation of the information that visitors provide.” (Quintarelli, 2005)

Folksonomies are not a top-down, hierarchical strategy, but rather a user-generated
classification through bottom-up consensus, they require people to associate keywords with the URL’s
of choice, and were born out of folk classification tools, such as Del.icio.us and 43things; social
bookmarking tools, and Flickr, a photo management, sharing tool, “has a similar system of free-form
tagging for photos that was adopted and modelled after Delicious” (Mathes, 2004). People using these
platforms tag their contents. Folksonomy is about the act of aggregation and not just the creation of
tags. Tagging gets better with scale, as long as at least one other person tags the way you would, you’ll
find it.
**Drawbacks**

“The problems with an uncontrolled vocabulary that is shared over an entire system leads to a number of problems and weaknesses in folksonomies.” (Mathes, 2004). This is not mentioned by Shirky who obviously is an advocate for this type of user classification.

As there is no controlled vocabulary, the terms can be ambiguous; people can use the same tag in different ways, for example searching under the keyword “tiger” brings up in a Google search, both “tiger” the animal, and “tiger” the Mac OSX Tiger. Similarly with lack of synonym control, different tags can be used for the same thing, for example, the Macintosh computer can be tagged as “Mac”, “apple”, “osx” and even “macosx”, found on a search in del.icio.us. To add to the vocabulary problems, is the different languages that are used throughout the world which may not have the same understanding or the same meanings. It can be argued that latest software releases have a correlation feature that, given a tag, shows related tags (Quintarelli, 2005). Folksonomies have a low search ability, great for serendipity, pot luck, and browsing, but not for specific items searches, but this too can be argued is not their purpose.

**World view, versus, world fitting into classification**

Shirky finishes his paper by stating that these systems benefit us because they don’t recreate the structured, hierarchical categorization so often forced onto us by our physical systems. “Does the world make sense or do we make sense of the world.” He argues that if we believe that the world makes sense and someone else is telling us that the world according to them is different to they way you see it, does it then make you wrong.

Mathes points out that “a folksonomy represents simultaneously the best and worst in the organization of information, its uncontrolled nature is fundamentally chaotic and suffers from problems of imprecision and ambiguity that well developed controlled vocabularies and name authorities effectively ameliorate.” He also points out that this very nature is also its best feature, the free form tagging that encourage users to organise the information according to the way that they would use that information themselves, being responsive to the user needs and with natural language.

I would rather have a compromise of the two systems, the new approach to classifying digital objects, and combine it with the traditional hierarchical classification, where users could freely create, use or reject terms stored in a database, but that is controlled by a selected authority that helps control the vocabulary.
Bibliography


[http://www.oclc.org/dewey/](http://www.oclc.org/dewey/)

[http://www.loc.gov/cds/index.html](http://www.loc.gov/cds/index.html)

[http://www.google.com](http://www.google.com)

[http://del.icio.us](http://del.icio.us)
