## Legal research skills training

Induction sessions are offered to new students and cover library services and use of the main legal databases. Training in legal research skills is currently integrated into the first-year legal methods class through lectures, tutorials and an information-skills-based exercise. Further integration, including formal assessment and additional online support, is scheduled for the forthcoming academic year.

The library continues legal research skills training through the VLE-based Computing and Communication for Lawyers (CACL) class, which takes an information-literacy-focussed approach to finding and using legal materials. This class is formally assessed and compulsory for all LLB students. The law school's VLE is used by library staff to communicate with students. This allows answers to recurring queries to be posted and specialist guidance to be targeted at individual classes. A revised research skills training programme incorporating web-cast presentations and Informs tutorials is currently in development.

## Further information

**Library**: http://www.lib.strath.ac.uk/
**Email**: library@strath.ac.uk
**Law School**: http://www.law.strath.ac.uk/
**University of Strathclyde**: http://www.strath.ac.uk/

*Legal Information Management*, 9 (2009), pp. 53–56
© The British and Irish Association of Law Librarians Printed in the United Kingdom doi:10.1017/S1472669609000140

# [On Not] Taming the Information Wilderness

**Abstract:** In this article, Julie Jones looks at the past, present, and possible future of information organisation, and considers how legal research might be conducted in years to come as the internet and computer science advance. Information classification systems have historically had finite life spans and there is no reason to believe that the current crop is any different. With increasingly more information being born digital, new ways of thinking about organisation and searching methodologies are developing to handle the deluge.
**Keywords:** information retrieval; internet; classification

## Introduction

Melvil Dewey did it. John West did it. Herbert Putnam did it, with the help of Charles Cutter. They all tried to tame the wilderness, and the wilderness overgrew them. Now Tim Berners-Lee is trying to do it again, this time with a completely different tack.

## The analog age

Dewey created the Dewey Decimal System in 1876. He created his classification scheme that would organise all the world's knowledge into ten classes: computer science and information, philosophy and psychology, religion, social sciences, language, science, technology, arts and recreation, literature, and history and geography. Each of these was then subdivided into ten divisions, each of which was then sub-subdivided into ten sections. Perfection attempted in decimal simplicity.



*Julie Jones*

Twenty years later in 1897, Putnam and Cutter put their heads together to create the Library of Congress Classification System, which utilised the conveniently pre-existing western alphabet as a method of organising



https://doi.org/10.1017/S1472669609000140 Published online by Cambridge University Press

information. Twenty-six primary categories, A through to Z, would contain the world's knowledge here. Originally designed to organise only the Library of Congress holdings, it is now used by libraries around the world.

At the same time, John West created an avalanche of legal information by publishing case law as it had never been published before, then created the Digest system for classifying that information (technically, John A. Mallory is said to be the original designer of the system, though West always gets the credit). Choosing a scant seven domains into which he would pour the entire world of US legal knowledge as dispensed by US courts, he helped mould the minds of attorneys for generations. Based on Langdell's first year courses at Harvard Law School, the categories are persons, property, contracts, torts, crimes, remedies, and government.

There are others, like Paul Otlet and Henri la Fountaine's Universal Decimal Classification system, created around the end of the 19th century and used in many countries. Henry Bliss created his own system in the mid-20th century, popular in the UK. S.R. Ranganatham invented his Colon Classification system in the early 20th century, popular in India. However, the Dewey and LC systems remain the leaders in libraries around the world to this day. Likewise, many common law countries have their own methods of organising case law as well, typically begun around the same general time-frame. For example, Canada uses the Canadian Abridgement, begun in the early 19th century.

All these men worked at a time when information was exploding, thanks to advancing literacy rates, increasing speeds of communication, and a flourishing publishing industry. The wealth of information, popularity of study, ease of travel, and advances in science created a growing mass of information (sound familiar?). It instilled in these men the deepest need to tame the information, to organise the information, to place it into discrete silos of information, so that it might be easily retrievable, accessible, and knowable.

These systems worked for many years. For each, the categories expanded, sometimes elegantly, sometimes monstrously, to accommodate the future. Now, all three systems are approaching their breaking points, as knowledge, information, and culture have burst well beyond the bounds imagined in the late nineteenth century.

And why wouldn't this happen? Every effort at information production and classification has decayed and been constructed anew as culture and technology transform. The Mesopotamian scribes organised clay tablets by data marked on the tablet edges. The librarians of Alexandria arranged scrolls alphabetically. Christian historians catalogued manuscripts by canon tables. Mass-produced books spawned the Dewey and Library of Congress systems, just as mass-produced case law gave birth to the Digest. These are only the latest attempts in an endless series of methods by which information is produced, collected, organised, and disseminated.

## The digital age

Then the internet brought chaos, both wonderful and terrifying in its expansiveness. All information. All accessible. All unorganised. All the time. (Not really, as we know. In fact, this is a rather pernicious myth.). It is in the process of rendering the old information guards obsolete (or at least in need of major reconstructive surgery).

After thirty-some years with the internet, instead of silos, we now have interconnection. Instead of pre-ordained classifications, we have user created tags. Instead of expertise of the few, we have wisdom of the masses. This, of course, is Web 2.0, the Social Web. It willingly and gleefully eschews authority and hierarchy, cutting at the very hearts of Dewey, Putnam, and West. Simultaneously, it presents challenging issues for the law which relies explicitly and completely on authority and hierarchy.

In this context, tags act very much like individual subject headings, topic and key numbers, or headnotes, but without the over-arching classification systems. They are one step beyond keyword searches which correlate only to exact text matches, and similar to indexes. Tags allow users to summarise material presented in a word or two, without being tied to the exact language of the material. Unlike expertly created hierarchical taxonomies, these folksonomies are user-generated and anti-hierarchical.

In this social web, information becomes like Tolkien's Ents, conscious trees that have the freedom of movement. They are not contained by any external arrangement. They are connected to their peers, not with ropes that bind, but with filament like a spider's web: flexible, strong, unimpeded. As a cute woodland creature ambles by, say with a new idea, a single strand might be lifted off one Ent and transferred to a new seedling. The entire forest shimmers with intellectual vibrancy.

## Digital chaos

Some readers are no doubt thinking that some of the individual animals may just be wandering aimlessly. In fact, a good number could be hopelessly lost. Others might suffer from rabies or be bad actors. Some readers may think that the masses are unreliable. Majority rule is mob rule! While Web 2.0 certainly allows for rogue activity, experience thus far has demonstrated that the mob generally works to keep the imps in line. Rabid raccoons and bad acting bunnies wreak havoc and minions create and edit publicly available information, and the overall effect is still useful.

Furthermore, classification schemes are hardly without fault. As the products of humans, they come with biases and prejudices factory installed, perpetuated and magnified over the years, skewing information like a



https://doi.org/10.1017/S1472669609000140 Published online by Cambridge University Press

black hole warps space, time, and light. Additionally, such schemes are surprisingly resistant to change. No system of information management has ever been, nor will ever be, perfect.

Web 1.0 research using Westlaw and Lexis keeps all primary and secondary sources in their separate silos, each with their own unique classification or indexing systems. While the capability exists in some circumstances to search across databases, the wise researcher knows that, because of how different types of legal information are created, published, and made searchable, combined database searches of different types of primary and/or secondary materials can be less than useful. Moreover, many free online sources of primary materials live in the dark web, inaccessible through popular internet searches, remaining undiscovered by the unaware and requiring users to search each website separately.

## Chaos managed?

So how does Web 2.0 transfer into Legal Research 2.0? Or, more accurately, what does open source taggable mashed up legal research look like? It looks like web sites where the full text of all cases are available for free to anyone – a place where anyone can add tags to cases based on the legal, procedural, and factual issues at play. A place where others can cross-reference tags to make new connections between cases that had previously resided in separate silos. It is a place where any member of the public can access a nation's jurisprudence for free, with the benefit of intuitive subject guidance (a.k.a. tags), without the burden and cost of proprietary silos and hierarchical language (a.k.a. topics and key numbers, or headnotes). Case law tags will link to expert commentary, to open source legislative codes, to regulations, and to administrative decisions. It portends novel legal arguments as new correlations and discoveries are made. Meaningful access to the law is abundantly granted with the relational and ubiquitous tags.

A rosy picture, indeed. Possible? – perhaps. Likely? Well… the reality will certainly be more complex, with subscription services like Westlaw and Lexis thrown into the mix. Perhaps only barristers will be allowed to tag or comment on free or subscription sites, as tagging and other social web applications are particularly vulnerable to those with nefarious intentions. Furthermore, some Legal 2.0 visions have failed to get too far off the ground, as experts tend to prefer monetary compensation for their intellectual efforts over goodwill. However, Web 2.0 applications remain built upon Web 1.0 technology. Computers continue to retrieve masses of results that technically satisfy search queries, but fail to intelligently analyse the results, leaving that task for us.

So let's move further into the future into what has been dubbed the Semantic Web, or Web 3.0, and the ever-expanding vision of Tim Berners-Lee, the next would-be wilderness tamer. As director of the World Wide Web Consortium (W3C), Berners-Lee actively works to further deconstruct organisational silos to create more understandable, more contextual, more relational information, and to force machines to think for themselves, thereby enhancing human knowledge and experience.

Despite tagging and the advances of Web 2.0, information remains hostage to software applications, which are their own types of silos. Web 3.0 seeks to create a web of data, instead of a web of pages as is our current experience. Using a different breed of coding, semantic web applications will:

(1) Allow computers to extract data from multiple sources and
(2) Recombine it for maximum effect using sophisticated analytical programming.

Instead of working with Ents in their entirety, we will have their bark, branches and leaves. A facile example of this is offered from the W3C website: "I can see my bank statements on the web, and my photographs, and I can see my appointments in a calendar. But can I see my photos in a calendar to see what I was doing when I took them? Can I see bank statement lines in a calendar"? W3C Semantic Web Activity, http://www.w3.org/2001/sw/.

It is a step beyond federated searching, which is now common in many libraries, where computer software allows users to search a university's catalogue and multiple subscription databases simultaneously. It conveniently allows numerous information silos to be searched at a single time, instead of searching each silo independently. However, these systems are still constrained by "unthinking" machine applications. There is no analysis in the results provided by the computer. It is just a list.

The goal of semantic applications in making data machine-readable is to advance human knowledge and make our lives easier. An example of the latter might mean that Ms. Barrister could instruct her computer to quickly reserve a flight, book a hotel, charge her credit card, bill the client, update her calendar, text message her partner, and send her reminders without her doing much at all. In fact, many science fiction movies feature Web 3.0 characters. Just look at "HAL" in *A Space Odyssey* and "Computer" in *Star Trek*.

## Chaos reconsidered

Today, this prospect may seem both exhilarating and menacing, depending upon one's perspective. The technology could make our daily lives easier, although there are privacy issues, to be sure. For machines to know what we want, they have to know a heck of a lot about us. Add that to our seemingly innate fear of computers that can think for themselves. Science fiction plots are



https://doi.org/10.1017/S1472669609000140 Published online by Cambridge University Press

replete with thinking computers gaining control over humans and going berserk.

Moreover, we must wonder what outsourcing our thoughts will have on our own humanity, intelligence, and society. (See Clive Thompson, *Your Outboard Brain Knows All*, Wired Magazine, Oct. 2007). Plato's fictional King Thamus famously begrudged the writing of information, fearing the loss of wisdom as humans relinquished their status as keepers of knowledge to mere papyrus. These same worries continue today and morph as machines do likewise. When we rely on computers for the memory of our own lives, for our own thoughts, will we be wiser? Will Web 3.0 technology free us up to think grander thoughts and make discoveries never before imagined? Or will it simply make us dull? The answer for most will lie somewhere in the middle, with the occasional heights of brilliance and depths of idiocy.

In its nascent stage, scientists and corporations are already exploring semantic web applications for deep information mining. Searching across platforms, they strive for new discoveries using old data recombined and remixed. Just think: the cure for cancer may be sitting right this minute in separate bits of information, lounging listlessly in lonely data silos, waiting for someone to introduce them.

How might the legal world benefit from Web 3.0? Perhaps law firm librarians will search their internal document management systems, address books, conflict checks, credit reports, newspapers, and case law simultaneously to investigate potential new clients. Attorneys may search case law, scientific journals, expert testimony, and witness databases to ascertain current jurisprudence and scientific standards, and identify which specialist is available to testify for a particular trial. Its applications to electronic discovery are sure to be epic, as email, cell phones, documents, voice mail, calendars, spreadsheets, and other records of communication and behaviour are opened even wider to outside scrutiny. Because of the different parameters and techniques used by Web 3.0 applications, search results will (allegedly) be smarter and substantially more relevant as the machines think about what makes the data useful to the human.

Legal Research 3.0 may not entail selecting a database to search based on jurisdiction or constructing complex Boolean queries. Instead, researchers may simply describe in plain language their legal issue, a description of their case, and what they need to know. The computer will search the internet, including all subscription databases available to the researcher, and all internal databases available within the researcher's organisational network. It will consider case law, legal periodicals, tags, briefs, legislation, topics and key numbers, blawgs, regulations, LC subject headings, memos, and interdisciplinary journals, in framing and organising the results, and suggest other searches or legal concepts not yet considered. In time, machines may analyse the law itself.

## Contextual chaos

I expect that these examples will be laughably shortsighted in, say, about five minutes. I also expect that, as we blend increasingly into our virtual world, we will carry with us a combination of taxonomies, folksonomies, and whatever else the wilderness cultivates. Nevertheless, Web 4.0 will soon beam onto the scene with quantum computers and other technologies not yet dreamed of. Suffice it to say, the times they are a-changing. The days of silos are fast coming to a close. The wilderness of information will never be tamed. It is simply the nature of information to be wild. The future of information is not containment and classification. The future is contextual chaos and cogitative computers.

## Biography

Julie Jones is currently Head of Information Services & Lecturer in Law at Cornell Law Library in Ithaca, NY, where she teaches basic and advanced legal research and writes on legal information and technology issues. She received her J.D. from Northwestern University School of Law in 1999, and her M.L.I.S. from Dominican University Graduate School of Library and Information Science in 2004.

Julie M. Jones
Head of Information Services & Lecturer in Law
Cornell Law Library
340B Myron Taylor Hall
Ithaca, NY 14853
jmj45@cornell.edu



https://doi.org/10.1017/S1472669609000140 Published online by Cambridge University Press