Beyond tags: towards a richer folksonomic ecology

Jon Dron
Athabasca University

Over recent years there has been considerable growth in the use of social software, in which user-generated content plays a major role. This trend is popularly known as Web 2.0. Prominent among the countless thousands of Web 2.0 sites are social networking systems (e.g MySpace, Bebo, Facebook and LinkedIn) resource sharing systems (e.g. Flickr, YouTube and SlideShare) personal publication systems (notably blogs), collaborative publication/editing systems (e.g. Wikipedia, GoogleDocs, CMAP) and social bookmarking & site discovery systems (e.g. del.icio.us, Google Search and Technorati). Most social systems employ tags to build folksonomies of user generated metadata. A folksonomy offers many benefits, notably in creating a dynamically evolving bottom-up structure that is better adapted to the needs of the community that uses it than even the most successful conventional classification systems. However, once a community reaches any significant size, problems such as ambiguity, synonymy and homonymy reduce the value of the tag-based taxonomy until it becomes virtually meaningless. Solutions employing ontologies and judicious use of WordNet can help, but such top-down automation takes away some of the flexibility and close adaptation to community needs that are the key strengths of the social system and introduce structural constraints that may overlay different cultural or epistemological biases. Such methods also fail to address problems such as mischievous or malevolent content and tags that increasingly affect social environments. In this talk I will discuss alternative ways of maintaining the strengths of folksonomies without resorting to non-social top-down techniques and technologies. In particular, I will describe developments in multi-dimensional tagging and speculate on the value of hierarchical tagging (tagging of tags) as a means of coping with large-scale public folksonomies as well as other scale-related issues that affect social systems. I will go on to explore approaches to community policing (soft security) including the use of reputation systems and multi-dimensional ratings. Finally, I will identify key research agendas in this area for the near future, including those relating to trust & privacy, interoperability & aggregability, and the balance of top-down and bottom-up design.

Athabasca University

- fully distance university

- open

- diverse adult learners with mixed skillsets

Crowd teaching

- learners can help each other to find things to help them learn

- easy in small groups but doesn't scale to bigger networks

- as groups change so do needs - to learn is to change

note using crowds to teach is my interest – but the principles discussed here apply across multiple disciplines and interests

Some possible solutions

- simple search (but how do we know what is good? Google as a recommender system is OK, but it is not very sensitive to context)

- using categories (but how do we know what is good?)

- recommendations (but can't cope with too many of them - not good if the crowd is big)

- recommender systems (but we are all different and to learn is to change - this is not like movie recommendations where tastes stay roughly similar)

Recommendations for learning

not just *that* something is good but *how* and *why*?

- learners need to know not just that something is good but how and why
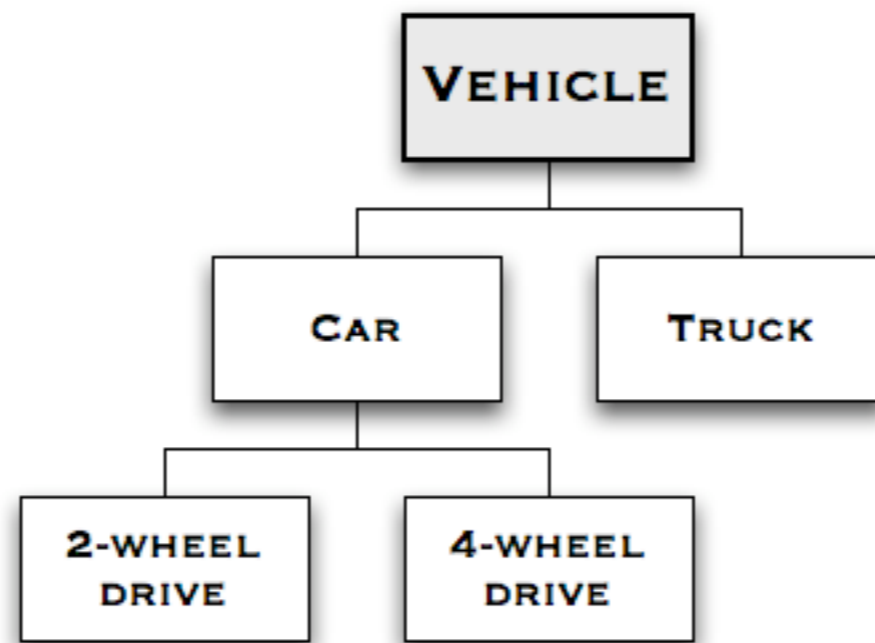
Folksonomies
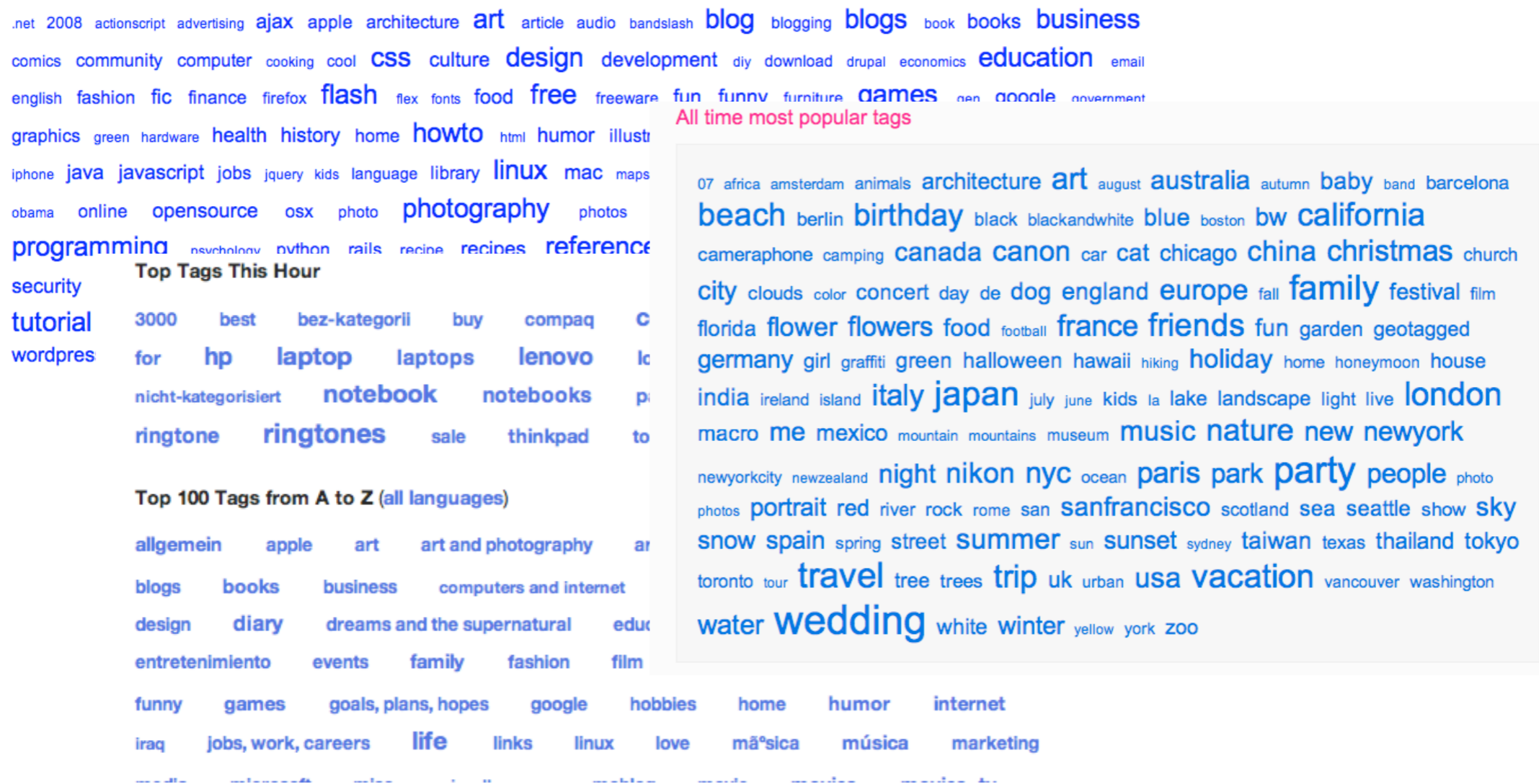
Definition of a tag – a label applied to something. Tag cloud contains weighted tags – people are about 3x more likely to click on a bigger word than a smaller one.

A particular form of recommender system – may use more or less complex algorithms that simply add or that consider similarity ot other users.

Typically tag clouds can show what a community finds interesting. A learner seeking guidance can follow the crowd, discovering what other people found useful.

- bottom-up classification

- Evolves to fit the community

the trouble with tags

- Ambiguity

- Synonymy

- Homonymy

- Diverse user needs and interests

- Stupid mobs vs wise crowds

- Multi-lingual issues

- Coarse grain - little context (why is it useful and how and where and for whom?)

Tags are useful, but can be misleading in many ways.
homonyms: e.g tag: a label, a lock of wool on a sheep, frequently repeated quotation, game, brand name, license plate, even argument for the existence of god (transendental argument for god).
Can have multiple tags meaning one thing (synonyms) – e.g. tag: label, topic, category, badge, stamp
Can mean similar things in different contexts and vice versa (e.g. html tags vs revision tag)
As sites get too big, meaning becomes diffuse and less relevant.

Some partial solutions

VEHICLE

CAR   TRUCK

2-WHEEL DRIVE   4-WHEEL DRIVE

```
dog, domestic dog, Canis familiaris
    => canine, canid
        => carnivore
            => placental, placental mammal, eutherian, eutherian mammal
                => mammal
                    => vertebrate, craniate
                        => chordate
                            => animal, animate being, beast, brute, creature, fauna
                                => ...
```

- rule-based AI

- Ontologies

- e.g. use of wordnet to match concepts

- but fundamental weakness in definitions that may not suit communities or learner needs and are inflexible

- can use collaborative filtering, which is good for matching preferences but learners, by definition, change. What worked before will not reliably predict what will work next.

- can create user models and adapt but, again, learners change

But...

- Diverse needs (top down not always great and we are largely replacing one teacher with another, albeit a bit enhanced by crowd intelligence)

- Shifting communities, shifting contexts

- Scale - size of community is a problem if too big or too small

- still no clues as to why a resource is useful unless explicitly programmed in

beyond simple tagging

- multi-dimensional tagging

- hierarchical tagging

- relationship tagging

- tag tagging

- tagged ratings

- multi-dimensional tagging - not just categories, but different kinds of expression about something

- hierarchical tagging - tree structures and organisations

- relationship tagging - how do you know someone, what do you think of them

- tag tagging - a method to accomplish all the above

- tagged ratings: values attached when tagging, such as ratings - to increase value and applicability of tags

multi-dimensional tagging

- e.g. CoFIND: pedagogic tags (qualities)

- e.g. SlashDot: message categories

- benefits to learners: not just what something is, but how something is good, why it is good, for whom...but in what context?

hierarchical tagging

- context is important - not just for orderliness and keeping lots of metadata under control, but also because evolution happens best in small isolated environments.

- need for diversity

- stimulation of creativity

- but should be small world networks - need weak and/or occasional links between isolated (parcellated) areas

- parcellating the landscape

- deep trees vs shallow structure (buckets of tags)

Hierarchies allow us to sub-divide a set of tags as different and new contexts evolve, but they increase complexity of interface (but can limit hierarchies by using buckets of tags), and there are some problems if similar tags needed across hierarchies and still difficulties with different meanings in different contexts. I have no solutions to that.

tagging relationships

- relationship tags - allow you to say more than 'this is my friend'

- note similarity to capabilities of RDF

- standards - XFN (XHTML Friends Network) apml (attention profile ML), FOAF (friend of a friend), ORMS (open relationship management system) SIOC (semantically interlinked online communities)

Tag tagging
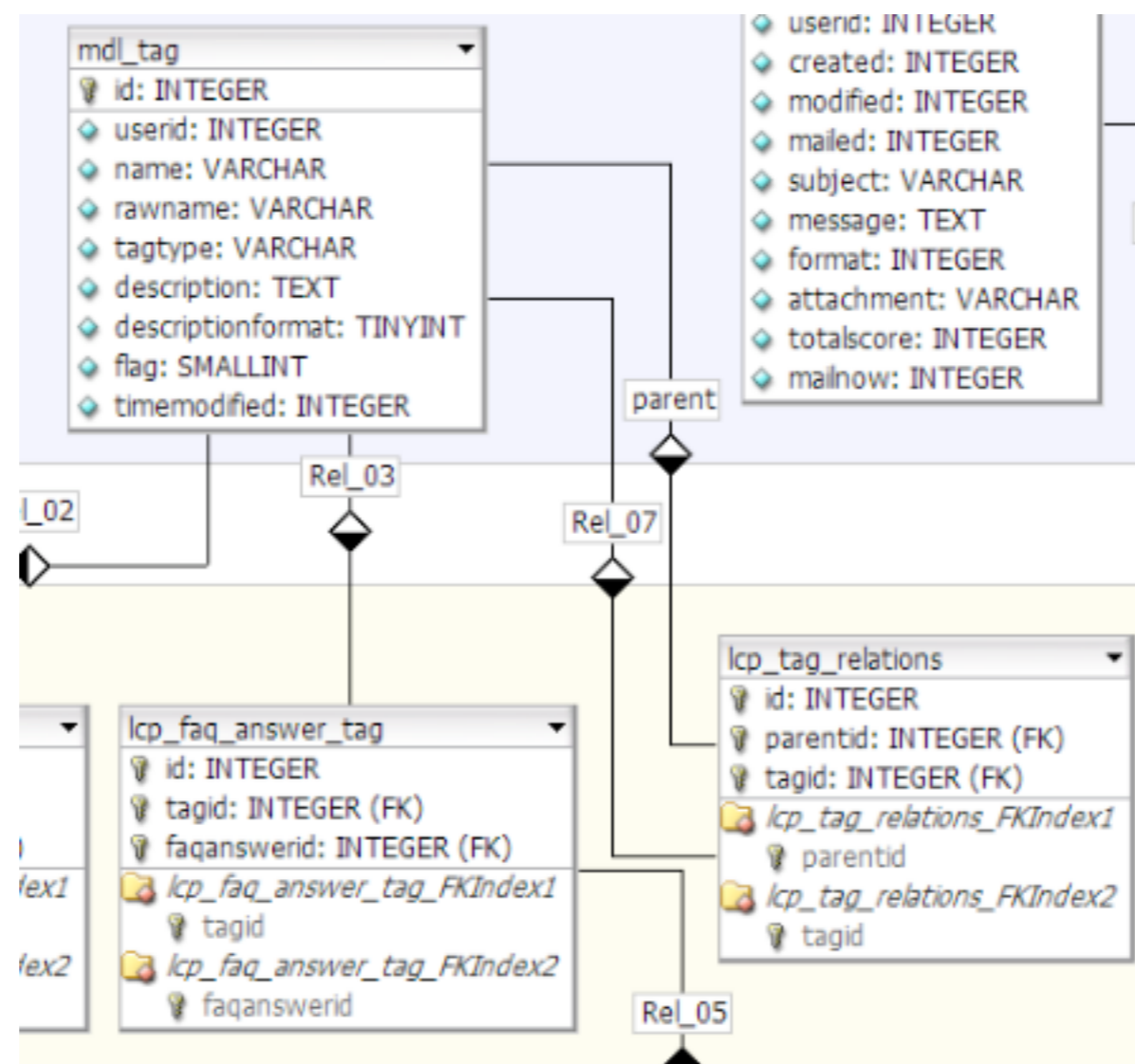
- **solution - you can tag the tags**

Tagging tags allows both hierarchies and multi-dimensional tags: tags that serve different purposes, such as pedagogical tags, relationship tags, hierarchical tags etc.

Tagged ratings

- not enough just to categorise

- ratings applied in the context of tags

- useful in many contexts - e.g. for people, for qualities/virtues

- can be explicit or implicit (e.g count of uses)

fitting it together

- with aforementioned principles learners can find stuff to help with their current problems, not bound to user model or teacher view. but this is not the whole story...

- aggregation integration parcellation - technical concerns - trust issues. How do we/should we transfer tags? (yes - we do so for RSS, for instance)

- transfer across domains, into and out of sub-domains - issues of relevance. No fixed ontology means that it may be hard to translate between different environments - meaning may change

- dataportability.org and beyond

trust and privacy

- how much do I want to share?

- need for fine-grained control over what i reveal of myself to who and when
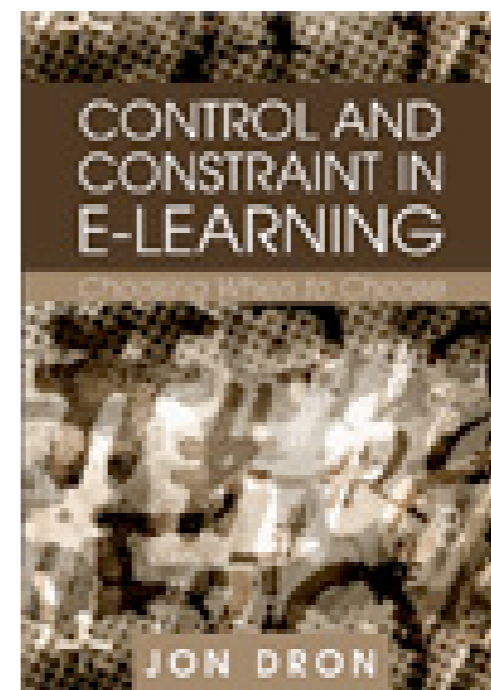
top-down vs. bottom-up

- structured metadata vs folksonomies

- jungles vs farms/gardens (note - Encyclopedia Britannica is a building/city or stuffed animals, Wikipedia is a farm, Flickr is wilder farm, del.icio.us is a jungle)

- role of the teacher as gardener/builder/farmer/guide/fellow-traveller

thank you

- jond@athabascau.ca

- Read the book...
  Control and Constraint in E-Learning:
  Choosing When to Choose

- jond@athabascau.ca

- Read the book...
  Control and Constraint in E-Learning: Choosing When
  to Choose
  IGI