Taxonomy challenges in digital publishing

Niké Brown
Enrichment Capability Specialist, John Wiley and Sons
My background...

- Honours degree in Information Management and Publishing
- Temp to permanent role at Croner Publications Ltd after graduation
  - Product development role for first content published on CD-ROM
  - Moved on to content management and thesaurus management roles
  - Manager of the Croner-I content platform
  - Spent three years as Content Architect leading a team of developers and content specialists
Wolters Kluwer UK and Wiley

- **Wolters Kluwer** are a global publishing company
- Reference publisher in finance, business and compliance, and healthcare
- Croner and CCH publishing houses formed WK UK through acquisition

- **Wiley** are a global academic publishing company
- Academic journal and scholarly research publisher
WK content online

- First attempt at creating an online-only product – disastrous...
- Croner-i: content created for online-only publishing – ahead of its time
  - ‘Smart’ content
  - XML
  - Thesaurus for classification metadata
  - Un-siloed content in contrast to books, etc
- Comparison with major competitor
  - ‘Books on screen’ – right down to the emulation of a ‘page’ flipping over
CHALLENGE #1: metadata and silos

- Pros of un-siloed metadata
  - Reuse of content
  - Flexibility for content configuration and online product development
  - Relating content previously buried
  - Maximising content assets
  - Enhanced user / customer experience
  - New revenue stream with multiple options to grow

- Cons of getting to un-siloed metadata
  - Cost
  - Effort
  - Resistance to change – new ways of working
Croner-I: metadata generated related content
Croner-i: metadata generated related content

Although stress isn’t a disability of itself, and so not covered by the Equality Act, it can be closely linked to other disabilities such as mental ill health and so should not be ignored by employers.

To this end, employers may want to use Stress Awareness Week as the catalyst for real action on their part to identify whether employees in their company are suffering from stress and, if they are, the potential causes of this.

It should be remembered that stress can lead to issues with productivity and performance due to related absences, potentially depriving a company of otherwise valuable members of staff.

Employees should feel safe and supported at work, and exploring ways that can prevent or work against stress before it gets too significant can help to combat these issues. As the Society says, proactivity is the key. Prevention is better than cure.

Last reviewed 29 March 2019
By contrast…
Wiley content online

- All journals and most reference works are on the Wiley Online Library
- Societies are entitled to have a Hub built by Wiley for their content, if they wish
  - Benefits of the Hub include enrichment
  - Content is ‘enriched’ with either an existing taxonomy, or a custom-built taxonomy
Different taxonomic approaches

Wolters Kluwer

- One (beautifully formed) thesaurus covering eight main market areas
- Active use of related terms
- Embedded as part of the Content Pipeline
  - Product builds would fail if content was not classified

Wiley

- Almost 200 taxonomies
- Currently, little reuse among content domains
- Audit of domains required
- Not part of the Content Pipeline (yet)
- Taxonomies come in many and various forms...
CHALLENGE #2: can you have too many taxonomies?

- One taxonomy or many?
- Croner went for one to cover all domains
- Wiley have many
- Software used
  - Concept schemes – can different projects or taxonomies be linked?
  - How are concept schemes treated? Wiley’s software treats concept schemes quite differently from Croner’s, which was different again from the original thesaurus management software used
  - Influences how you approach the formation of your thesaurus
CHALLENGE #3: understanding and expectations

- Internal resistance
  - New ways of working often required
  - "It’s not broken, why fix it?"
  - “Why can’t X do it – I’m too busy”

- Working with SMEs
  - Often have a mixed understanding of what’s required from them
  - The kitchen sink HAS to be included!
  - Anxiety that something essential won’t get covered

- Business expectations and views of enrichment
  - “It’s just a mechanical tool, isn’t it?”
  - “What’s my ROI?”
  - “What do you mean, it might never be finished??!”
Reactive or proactive?

- To quote Henry Ford: “If I’d asked people what they wanted, they’d have said, ‘A faster horse!’”
- It’s not always easy for non-taxonomists to see the benefit of content classification
  - Have to accept that some people will never see the point in taxonomies
- Great advantage in doing the work before the business realises it needs it
- Look for opportunities to enrich content and display the power of metadata
  - More persuasive than discussions, etc
Future challenges

- Embed taxonomy application into the content pipeline
- Promote understanding and enthusiasm for taxonomic classification
- Explore machine learning to build taxonomies
  - Content mining and entity extraction
- Expansion of taxonomy features on front end
- Development of ontologies
“At the end of the day, most of our categorisation decisions are pragmatic ones, which is why so many information scientists need to forget a lot of their training if they are to design knowledge taxonomies that work in practice.”

Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness
Patrick Lambe, 2007
Thanks for listening!

Questions?