

# Spokes in the Wheel: The Many Facets of Working with an Automatic Indexing System

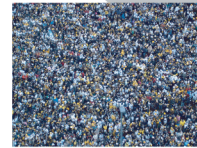


## Creativity



New ways of solving problems; new ideas for development. It's much more than simply being immersed in detailed text analysis

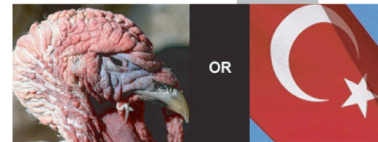
## Multi-Disciplinary Working



We work closely with developers, producers, designers and editorial staff. IAs are the acknowledged metadata experts who put users at the centre of everything they do

## Love of Language

We train the system to process linguistic ambiguities that humans wouldn't skip a beat over. We can tell these Turkeys apart by building sophisticated rules around context and other keywords in text

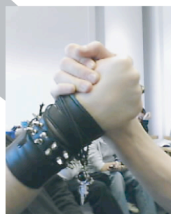


## Information Organization



Underpinning the indexing system are six controlled vocabularies arranged in facets called: Subject, Proper Name, Location, Time Period, Brand and Audience. A lot of thought goes into managing new and existing terms in those taxonomies and authority lists

## Collaboration with Users



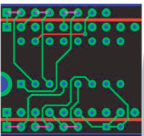
Users (and potential users) of the system are scattered across different projects and locations and have differing skills levels. We provide support materials, feedback and training

## Championing the System



It's important to evangelise the benefits of automated indexing, at a time when many big BBC projects require metadata of some description. We share knowledge with other departments and let them experiment with the system through an API

## Technical Awareness



It's helpful to be a little bit of a geek sometimes – understanding XML, say, or analysing indexing algorithms to improve them

## General Knowledge



Otherwise known as knowing a little about a lot of things. Our controlled vocabularies cover a huge variety of bbc.co.uk content – everything from the traditional Gloucestershire cheeserolling event to the latest big TV dramas

## About the System

The metadata system was developed by the Content Management Culture (CMC) team of BBC New Media. CMC builds and deploys content management solutions, based on Documentum, for clients around the BBC. Successful implementations have been English Regions and Children's BBC

Users must apply at least one metadata term to every object created in the content management system. Metadata is picked from controlled vocabularies, or users can submit new terms for review by IAs. The key uses for metadata are to support internal and external search engines, and to facilitate content reuse, sharing, aggregation and syndication

The indexing system is rules-based, meaning that IAs have total and immediate control over how terms are triggered when text is parsed through the system. Other approaches such as Bayesian were considered, but only a rules-based methodology allowed such flexibility and human control. These were important factors considering the breadth of content likely to be coming through the CMS (ruling out using training sets of documents to prime the system beforehand)

The indexing system is part of the Documentum content management solution, and is called Content Intelligence Services (CIS). The rules management tool is Domain Map Editor (DME). It is a controlled vocabulary management tool, and allows IAs to build indexing rules around boolean logic using relevant keywords, and ascribing confidence rankings to words and phrases