How the BBC is approaching metadata quality measurement

Jeremy Tarling, Lead Data Governance Specialist
BBC Design + Engineering
How the BBC is approaching metadata quality measurement

- BBC content tagging 101
- Problems with the current approach
- Building better metadata - the Common Metadata project
- Introducing machine tagging
- Metadata quality measurement
BBC content tagging 101: Content <- about -> Thing
BBC content tagging 101

Journalism

News, Sport

- Human tagging by subject
- > 2,000,000 tagged articles
- > 70,000 ‘Thing’ tags
- > 5,000 topic pages
BBC content tagging 101: journalism
BBC content tagging 101: journalism

**NEWS**

**Boris Johnson**

00:01

‘Big gap’ remains as Brexit deal talks to continue

The future of Northern Ireland’s border dominates, while Parliament reconvenes for the Queen’s Speech.

Read more ▶

20:01, 13 Oct

PM says ‘significant’ work to do on Brexit deal

---

**SPORT**

**Manchester United**

**Inter’s Sanchez Injured on Chile duty**

World Cup winner Schweinsteiger retires

How bad could it get for Man Utd?

Man Utd displays unacceptable - De Gea

Newcastle win to deepen Man Utd worries

---

**NEWS**

**Wolverhampton**

England, United Kingdom

Set as your local news? Yes

+5 miles +10 miles +20 miles +30 miles

LIVE Midlands Live: Breaking news and local stories

The pain that Wolverhampton stab victim’s killers walk free

4 Four injured in ‘mass brawl’ at wedding

Pedestrian hit by 4x4 dies

---

**SPORT**

**Athletics**

**Kosgei shatters women’s marathon record**

Watch Kosgei smash Radcliffe's marathon world record

Farah eighth as Cherono wins Chicago Marathon

Kipchoge breaks two-hour marathon mark

Kipchoge expects others to repeat marathon feat

Reaction after Kipchoge runs
BBC content tagging 101: journalism

Person

Organisation

Place

Theme
### BBC content tagging 101

#### Journalism

<table>
<thead>
<tr>
<th>News, Sport</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Human tagging by subject</td>
</tr>
<tr>
<td>• &gt; 2,000,000 tagged articles</td>
</tr>
<tr>
<td>• &gt; 10,000 topic pages</td>
</tr>
</tbody>
</table>

#### Programmes

<table>
<thead>
<tr>
<th>iPlayer, Sounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>• No tagging by subject (except R4)</td>
</tr>
<tr>
<td>• Human curation of content by genre and format</td>
</tr>
</tbody>
</table>
BBC content tagging 101: iPlayer and Sounds
## BBC content tagging 101

<table>
<thead>
<tr>
<th>Journalism</th>
<th>Programmes</th>
<th>the rest...</th>
</tr>
</thead>
<tbody>
<tr>
<td>News, Sport</td>
<td>iPlayer, Sounds</td>
<td>Bitesize, Recipes, etc</td>
</tr>
<tr>
<td>• Human tagging by subject</td>
<td>• No tagging by subject</td>
<td>• Bespoke domain models</td>
</tr>
<tr>
<td>• &gt; 2,000,000 tagged articles</td>
<td>• Human curation of content by genre and format</td>
<td>• Human tagging and curation</td>
</tr>
<tr>
<td>• &gt; 10,000 topic pages</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
BBC content tagging 101: Bitesize, Recipes, etc...
Problems with the current approach
Process, accuracy, quality and scale
Poor tagging process

embedded in journalism workflows 😊
but...

70,000 tags to choose from 😞
no taxonomic hierarchy to filter on 😞
easy to mis-tag, no second pair of eyes 😞

some tags are ambiguous 😞
Poor tagging accuracy

It can be hard for a human to deal with tag ambiguity

For example, Sport have topic pages for all olympic disciplines, including shooting...
Poor tagging accuracy

It can be hard for a human to deal with tag ambiguity

For example, Sport have topic pages for all olympic disciplines, including shooting...

But shooting can sometimes happen outside of sport...
Poor tagging accuracy
Poor tagging accuracy

Place names can be particularly problematic, especially when there are sports teams with the same name.
Poor tagging accuracy

Place names can be particularly problematic, especially when there are sports teams with the same name.

When a big News story is associated with a place it’s all too easy to mistake the team for the location.
Poor tagging accuracy

Rotherham United

'They threatened to rape my mum'

Some of the 1,400 children abused in Rotherham between 1997 and 2013 speak about what happened to them.
Poor tagging quality

5% of our tags are broad and cover 85% of our content

95% of our tags are specific and cover 15% of our content
Hardly any TV and radio tagging

Genre: 20010 (comedy)

Genre: 20010 (comedy)

according to our TV metadata these programmes are the same...
Building better metadata: the Common Metadata project
Building better metadata: defining common metadata

generalised schema for any BBC content object

new controlled vocabularies for editorial tone and intended audience

building in automation from the start by modelling **provenance**
Building better metadata: more than just ‘about’ tagging

There’s a strong appetite in TV and Radio teams for better metadata, especially around editorial tone and audience behaviour.

There’s a willingness to experiment with automatic classification by these dimensions, but with human curation on top... for now.
Building better metadata: editorial tone

working with commissioning experts and UX to define metadata to represent the editorial of a piece of content
Building better metadata: intended audience

working with UX and M&A experts to define metadata to represent the intended audience of a piece of content

<table>
<thead>
<tr>
<th>preferredLabel</th>
<th>disambiguationHint</th>
<th>type(s)</th>
<th>URI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Intended audience: gender</td>
<td>core:Thing</td>
<td><a href="http://www.bbc.co.uk/things/19eb702ec-0b30-4d12-869-859ad6634f6d">URI</a></td>
</tr>
<tr>
<td>Female</td>
<td>Intended audience: gender</td>
<td>core:Thing</td>
<td><a href="http://www.bbc.co.uk/things/3aa70400-7571-421e-97e6-91a9876281a9f6d">URI</a></td>
</tr>
<tr>
<td>Non-binary</td>
<td>Intended audience: Gender</td>
<td>core:Thing</td>
<td><a href="http://www.bbc.co.uk/things/ed90b445-1fd4-468a-9952-208812a603ef6d">URI</a></td>
</tr>
<tr>
<td>ABC1</td>
<td>Intended audience: situation</td>
<td>core:Thing</td>
<td><a href="http://www.bbc.co.uk/things/8c31e6e6-1159-43b0-9dc9-73e5a27eda9ef6d">URI</a></td>
</tr>
<tr>
<td>C2DE</td>
<td>Intended audience: Situation</td>
<td>core:Thing</td>
<td><a href="http://www.bbc.co.uk/things/2fc5c337-b0d7-494e-a852-1d8e88f2d039f6d">URI</a></td>
</tr>
<tr>
<td>Under-35</td>
<td>Intended audience: Age</td>
<td>core:Thing</td>
<td><a href="http://www.bbc.co.uk/things/376c1218-21d8-498a-a66b-025dc72bc8fd6d">URI</a></td>
</tr>
<tr>
<td>Over-35</td>
<td>Intended audience: Age</td>
<td>core:Thing</td>
<td><a href="http://www.bbc.co.uk/things/e54e8a5a-9fde-411f-a0c3-7517736e058f6d">URI</a></td>
</tr>
<tr>
<td>Inform me</td>
<td>Intended audience: motivation</td>
<td>core:Thing</td>
<td><a href="http://www.bbc.co.uk/things/bf928a3-b38d-4647-924e-c03bdc29174f6d">URI</a></td>
</tr>
<tr>
<td>Educate me</td>
<td>Intended audience: target</td>
<td>core:Thing</td>
<td><a href="http://www.bbc.co.uk/things/b1e660eb-5aad-4d30-b2c6-26acdc29174f6d">URI</a></td>
</tr>
</tbody>
</table>
Audience behaviours as metadata - “News needs”

OUTWARD
KEEP ME ON TREND
INSPIRE ME

FUNCTIONAL
UPDATE ME
AMUSE ME
EMOTIONAL

GIVE ME PERSPECTIVE
EDUCATE ME

INWARD
Building better tooling: Passport Control

Passport Control is the user interface to CRUD Passports

Useable as a standable tool or (better) embedded as an activity in an existing content production workflow
Improving metadata governance

Over the past decade we’ve evolved a vocabulary of 70,000 Things that can be used as tags.

This makes it much harder for a human to tag a piece of content with (for example) Boris Johnson.

Machine tagging can help, but really we need better structure and governance of our tags.
Introducing machine tagging
Starfruit, pace/valence and Vox
First steps in automation: about and mentions

**Starfruit** - developed by BBC R&D

uses a standard machine learning approach to text classification (Multi-label Classification)

trained on a subset of the 70,000 tags that have been used at least 10 times to describe journalism content over the past few years

can be used to suggest tags to a journalist
First steps in automation: Starfruit “aboutness”

<table>
<thead>
<tr>
<th>Preferred Label</th>
<th>URI</th>
<th>Disambiguation Hint</th>
<th>Score</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Politics</td>
<td><a href="http://www.bbc.co.uk/things/75612fa6-147c-4a43-97f4-6f70d9c0ed3#id">http://www.bbc.co.uk/things/75612fa6-147c-4a43-97f4-6f70d9c0ed3#id</a></td>
<td>Governance of states</td>
<td>0.8580</td>
<td>tagging:TagConcept, core:Theme, core:Thing</td>
</tr>
<tr>
<td>Boris Johnson</td>
<td><a href="http://www.bbc.co.uk/things/6f79dd85-35b5-40cd-1366-1e152d69fba#id">http://www.bbc.co.uk/things/6f79dd85-35b5-40cd-1366-1e152d69fba#id</a></td>
<td>UK politician</td>
<td>0.7953</td>
<td>tagging:TagConcept, core:Person, core:Thing</td>
</tr>
<tr>
<td>Jeremy Corbyn</td>
<td><a href="http://www.bbc.co.uk/things/71e97cc0-8bec-416b-8210-6a6bb8571f1#id">http://www.bbc.co.uk/things/71e97cc0-8bec-416b-8210-6a6bb8571f1#id</a></td>
<td>UK politician</td>
<td>0.6565</td>
<td>tagging:TagConcept, core:Person, core:Thing</td>
</tr>
<tr>
<td>UK Parliament</td>
<td><a href="http://www.bbc.co.uk/things/1d411bc6-c505-4437-8139-7536c7697f1#id">http://www.bbc.co.uk/things/1d411bc6-c505-4437-8139-7536c7697f1#id</a></td>
<td>Legislative body of the UK</td>
<td>0.5871</td>
<td>tagging:TagConcept, core:Thing, core:Organisation</td>
</tr>
<tr>
<td>Unlawful Parliament suspension</td>
<td><a href="http://www.bbc.co.uk/things/9f4db5957-ac41-4e5a-829d-36a173a4b15b9#id">http://www.bbc.co.uk/things/9f4db5957-ac41-4e5a-829d-36a173a4b15b9#id</a></td>
<td>Suspension of Parliament in the run up to Brexit, 2019</td>
<td>0.5501</td>
<td>tagging:TagConcept, core:Thing, core:Event</td>
</tr>
<tr>
<td>Brexit</td>
<td><a href="http://www.bbc.co.uk/things/fd4293fd-f83-400e-8bc6-89ec01930fcf#id">http://www.bbc.co.uk/things/fd4293fd-f83-400e-8bc6-89ec01930fcf#id</a></td>
<td>UK departure from the EU</td>
<td>0.5337</td>
<td>core:Thing, tagging:TagConcept, core:Event</td>
</tr>
</tbody>
</table>

Boris Johnson’s government has set out “ambitious” policies on crime, health, the environment and Brexit in a Queen’s Speech opposition parties have dismissed as a ‘manifesto’. Plans for tougher sentences for violent offenders, legal targets for cutting plastic pollution are bills set out at Parliament’s State Opening.

The BBC’s Laura Kuenssberg said it was a ‘shopping list’.
“Our political editor said the PM was keen to focus on "bread and butter issues" like investment in schools and the NHS, or coming up with, at long last, a new way of funding care for the elderly.”
First steps in automation: auto-tagging works for journalism

R&D’s Starfruit is building confidence about machine classification in our editorial community.

So far we are only using it for suggesting tags, but we are pushing for full auto-tagging.

Content creators worry about automation of these editorial tasks, both for accuracy and for their job security.

Last month, Sports Direct complained that its rival JD Sports' planned £90m takeover of Footasylum could reduce Mr Ashley’s access to the top brands.

"Sports Direct has consistently aimed to provide the widest range of products at attractive prices and will continue to work constructively with all its suppliers to enhance its product offering for the benefit of consumers," the company said.

The BBC has asked Nike and Adidas for comment. In response to the Sunday Times story, Nike told the newspaper that it made adjustments to its sales channels from "time to time" to "optimise distribution".

Related Topics

<table>
<thead>
<tr>
<th>Sports Direct</th>
<th>Retailing</th>
<th>Mike Ashley</th>
<th>Companies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adidas</td>
<td>Nike, Inc.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
First steps in automation: editorial tone

BBC R&D are developing an automated editorial tone classifier for TV drama segments based on *valence* and *arousal*.

"**Valence**, as used in psychology, especially in discussing emotions, means the intrinsic attractiveness/"good"-ness (positive valence) or averseness/"bad"-ness (negative valence) of an event, object, or situation."

"**Arousal** involves activation of the ascending reticular activating system in the brain, which mediates wakefulness, the autonomic nervous system, and the endocrine system, leading to increased heart rate and blood pressure and a condition of sensory alertness, mobility, and readiness to respond."
First steps in automation: editorial tone

initial model built on human classification of comedy and drama clips

small controlled vocabulary representing the vectors - “slow/fast”, “exciting/boring”, “happy/sad”, etc.

machine classification based on existing model then retained
First steps in automation: editorial tone
Metadata quality measurement
Completeness, Accuracy, Provenance
Metadata quality measurement

Metadata quality measurement is well documented in library and information science.

How might we develop a quality metric for the BBC’s descriptive content metadata?

- completeness
- relevancy
- conformance to expectation
- accuracy
- timeliness
- accessibility
- provenance
- fitness for purpose
Metadata quality measurement: Completeness

Given an agreed set of metadata properties, is there any metadata present for these fields?

Are all fields equal in importance?

Across all domains/products?
## Metadata quality measurement: Completeness

<table>
<thead>
<tr>
<th>property</th>
<th>News articles</th>
<th>Sounds music mixes</th>
</tr>
</thead>
<tbody>
<tr>
<td>about</td>
<td>90%</td>
<td>10%</td>
</tr>
<tr>
<td>editorial tone</td>
<td>10%</td>
<td>80%</td>
</tr>
<tr>
<td>Intended audience</td>
<td>20%</td>
<td>30%</td>
</tr>
</tbody>
</table>
Metadata quality measurement: Provenance

Initially we are dealing with two provenances, human vs machine

But...

- Human tagging quality will be domain-specific and affected by experience and knowledge

- Machine tagging quality will be model-specific and only as good as the quality and timeliness of the training set
# Metadata quality measurement: Provenance

<table>
<thead>
<tr>
<th>source</th>
<th>News articles</th>
<th>Sounds music mixes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journalist</td>
<td>90%</td>
<td>40%</td>
</tr>
<tr>
<td>Music producer</td>
<td>50%</td>
<td>90%</td>
</tr>
<tr>
<td>Starfruit (model 94)</td>
<td>75%</td>
<td>25%</td>
</tr>
</tbody>
</table>
Metadata quality measurement: Accuracy

Accuracy is calculated as a function of provenance and completeness at the individual property level.

“User X with provenance score 25% for news articles says https://www.bbc.co.uk/news/uk-england-derbyshire-51611555 is → about → TripAdvisor, NatWest bank, Ilkeston”

“User Y with provenance score 75% for news articles says https://www.bbc.co.uk/news/uk-england-derbyshire-51611555 is → about → Ilkeston → tone → funny”
Metadata quality measurement: CAP scores

We take **Curation**, **Accuracy** and **Provenance** together to generate a ‘**CAP score**’ for a given set of metadata assertions

\[
\text{CAP} = \frac{(\text{threshold} + \text{SUM(relevant assertion confidence scores)})}{(\text{number of relevant assertions} + 1)}
\]

(relevant assertions = assertions that support the value, e.g. assertions that would add the same About tag A to the metadata set)
Metadata quality measurement: CAP scores

Event store records

<table>
<thead>
<tr>
<th>Data source</th>
<th>Sensitivity</th>
<th>About tag A</th>
<th>About tag B</th>
<th>About tag C</th>
</tr>
</thead>
<tbody>
<tr>
<td>User 1</td>
<td>Sensitive</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>User 2</td>
<td>Not sensitive</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

individual values’ CAP scores

- $\text{CAP(Sensitivity)} = \frac{0.85}{2} = 0.425$
- $\text{CAP(About tag A)} = \frac{0.85}{2} = 0.425$
- $\text{CAP(About tag B)} = \frac{(0.85 - 0.75)}{2} = 0.8$
- $\text{CAP(About tag C)} = \frac{0.75}{2} = 0.375$

attributes’ CAP scores

- $\text{CAP(Sensitivity)} = 0.425$
- $\text{CAP(About)} = 0.425 + 0.8 + 0.375 = 1.6$

final CAP score

$\text{CAP} = 0.9 \times 0.425 + 0.8 \times 1.6 = 1.78$

User 1 (confidence: 85%)

User 2 (confidence: 75%)

Passport record

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>About taggings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitive</td>
<td>A, B, C</td>
</tr>
</tbody>
</table>
Thanks for listening!

Links:

BBC Things (tags): https://www.bbc.co.uk/things/

BBC Ontologies: https://www.bbc.co.uk/ontologies/

Starfruit: http://starfruit.virt.ch.bbc.co.uk/

Vox: http://vox.virt.ch.bbc.co.uk/

@jeremytarling
jeremy.tarling@bbc.co.uk