

Position Paper: W3C Workshop on Annotation: Semantic content aggregation and discovery powered by rich business domain annotations.

Submitted by: Jem Rayfield, Head of Solution Architecture, The Financial Times

The Financial Times (FT), publishes international business and economic news. The newspaper and online publications are published by Pearson PLC in London. The FT is considered the worlds most important business read with a readership across senior financial decision makers within the worlds largest financial institutions. The FT's audience is shifting from traditional print media towards online. The FT launched FT.com in 1996 and pioneered a metered paywall in 2002 and is one of the few UK online publishers to successfully operate a subscription model. In 2007 the FT launched an innovative HTML 5 mobile web app. Smartphone and tablets now drive 20% of traffic to FT.com. More than half of the FT's subscriptions are now digital.

The FT operates within a multiplatform landscape and must make its content universally accessible across all of its channels. To achieve this the FT annotates all of its content objects using a business domain centric semantic fingerprint. The FT stores its articles, blogs, images, videos and comments linked via annotation to a rich business/economic domain model.

The FTs online publishing stack uses a dynamic semantic architecture to underpin its formation. This linked data architectural approach makes use of RDF, inference and NLP to power annotation and its API powered online products.

The FT is particularly interested in standardised annotation ontologies and models:

- Standardised domain model annotation; the FT has curated its own annotation ontology to link Annotatable content objects to real-world concepts. This simple ontology has predicate equivalences to ontologies such as rNews and Schema.org (about/mentions/etc). The FT would like to move (contribute) to a standardised annotation ontology.
- Annotation Provenance; the FT extends its annotation ontology with provenance. Provenance information such as the author, published dates, system curator etc. The FT would like to move (contribute) to a standardized annotation provenance scheme.
- Position and offset vs embedded annotation; the FT has modelled annotation positional attribution using position and offset modelling. The FT would like to understand how others have implemented similar use cases.
- Time coding; the FT converts its audio and video soundtracks to text. The FT plans to use time coded annotations across its a/v. In addition it would like to fragment video/audio into clips using semantic annotation/fingerprinting. The FT would like to understand if there are any other use-cases and indeed models/ontologies used by other participants of the workshop? Are there any forming standards?

- Comment annotation; the FT currently allows users to comment on content objects. It would like to allow comments to be more granular and provide greater provenance. The FT wishes to understand workshop participant usecases, solutions and models.

The FT has a set of content, annotation and business domain ontologies which are beginning to power its online publishing stack. The FT would like to contribute to and understand emerging annotation ontology standards.