

DATA MODEL FOR EMBEDDING METADATA IN WEB DOCUMENTS BY RNEWS STANDARD

Olena Balalaieva

National University of Life and Environmental Sciences of Ukraine,
Heroiv Oborony Str. 15, Kyiv, 03041
olena.balalaieva@gmail.com

Abstract

The article deals with rNews – an approved standard for using semantic markup to annotate news-specific metadata in HTML documents. The relevance of the study is due to the objective need for theoretical justification and the specification of new communication models that have arisen with new channels of communication emergence, media digitalization and convergence. The analysis of rNews data model components are given in the paper. The experience of the world's leading agencies can be an important benchmark for implementation the rNews data model in domestic practice and studying the theoretical aspects of its use.

Keywords: data model, metadata, standard, attribute, news, mass communication, media

1. Introduction

Among the general trends in the development of modern media for 2020, experts note the expansion of the use of advanced technologies, the emergence of new information portals, technically more democratic information sharing systems, and increasing audience participation in media communications [1]. Transmediality and democratization of the content distribution and consumption system provide for the change of the characteristics of information as the basic category of media communications.

V. Stepanov considers the imperative of communicative behavior, which is ensured by verbal, reflective, virtual, polycodal and other characteristics as a basic standard-setting principle of modern mass communication. The researcher emphasizes that the multimedia potential of cyberspace promotes multi-channel and multi-code impact of information: through vision (video), hearing (audio), space (proxemics), time (chronemics), etc. The speed of multimedia information processing becomes a factor that affects the ability of a person to survive in modern conditions polycodal information environment [2].

These trends should be provided with adequate models of communication in the mass media. The journalist as a subject of communication has to create interaction with the audience on the content and technical levels through the creation of emotionally attractive polycodal virtual communication.

Researchers [3] emphasize that one of the means of increasing polycode, emotional attraction, simplifying the content usage and thereafter removing the barriers to involving the audience to communication, is multimedia, the introduction of which is due to new abilities of modern media as a result of digitization.

If the investigation object is being developed, the modeling method is often used in the scientific practice, because the model presents the properties, connections, tendencies of the systems and processes under study.

The purpose of this article is to analyze the components of the data model in news media by the standard of the International Press Telecommunications Council.

2. Analysis of rNews data model components

The International Press Telecommunications Council (IPTC) is an organization that brings together the world's largest news agencies, deals with the development and implementation of technical standards for the exchange of information between content providers and consumers. Today it has more than 50 companies and news industry organizations, including BBC, Thomson Reuters, The New York Times and other are member of IPTC.

The goal of IPTC is to create and optimize technical standards for the exchange of news. The IPTC mission is to simplify the distribution of information.

As stated on the official website, IPTC creates and supports sets of concepts that can be assigned as metadata values for objects such as text, photographs, graphic images, audio and video files, and streams. This allows for consistent coding of news metadata for a certain period For the news industry, this is a strict requirement that allows you to define news content and apply metadata [4].

One of the standards developed by IPTC is rNews. Today rNews is an approved standard for using semantic markup to annotate news-specific metadata in HTML documents.

The initiator of the standard development was The New York Times, which officially asked IPTC at the end of 2010 to take the initiative in developing a standardized set of classes and attributes for embedding publication metadata in HTML documents. Because, as it shown in Figure 1, some story components which are obvious to a person are not obvious to a machine parsing HTML.



Figure 1. Differences between human reception and machine parsing HTML of websites (from official IPTC website)

Modern websites have a three-tiered architecture: a data tier (consists of one or more databases), presentation tier (HTML document and accompanying resources (scripts, images, etc.) delivered by the logic to the requesting user's browser) and a logic tier, (software that reads from data tier and outputs the presentation tier). So content is well structured on data tier, but non on presentation tier. There is a problem of structured data at different levels.

The standard developed to solve this problem «rNews specifies the terminology and data model required to embed news specific metadata into HTML documents» [4].

Let's look at the rNews data model, built on the basis of observations, theoretical generalization and improvement of the experience of leading news agencies (Figure 2).

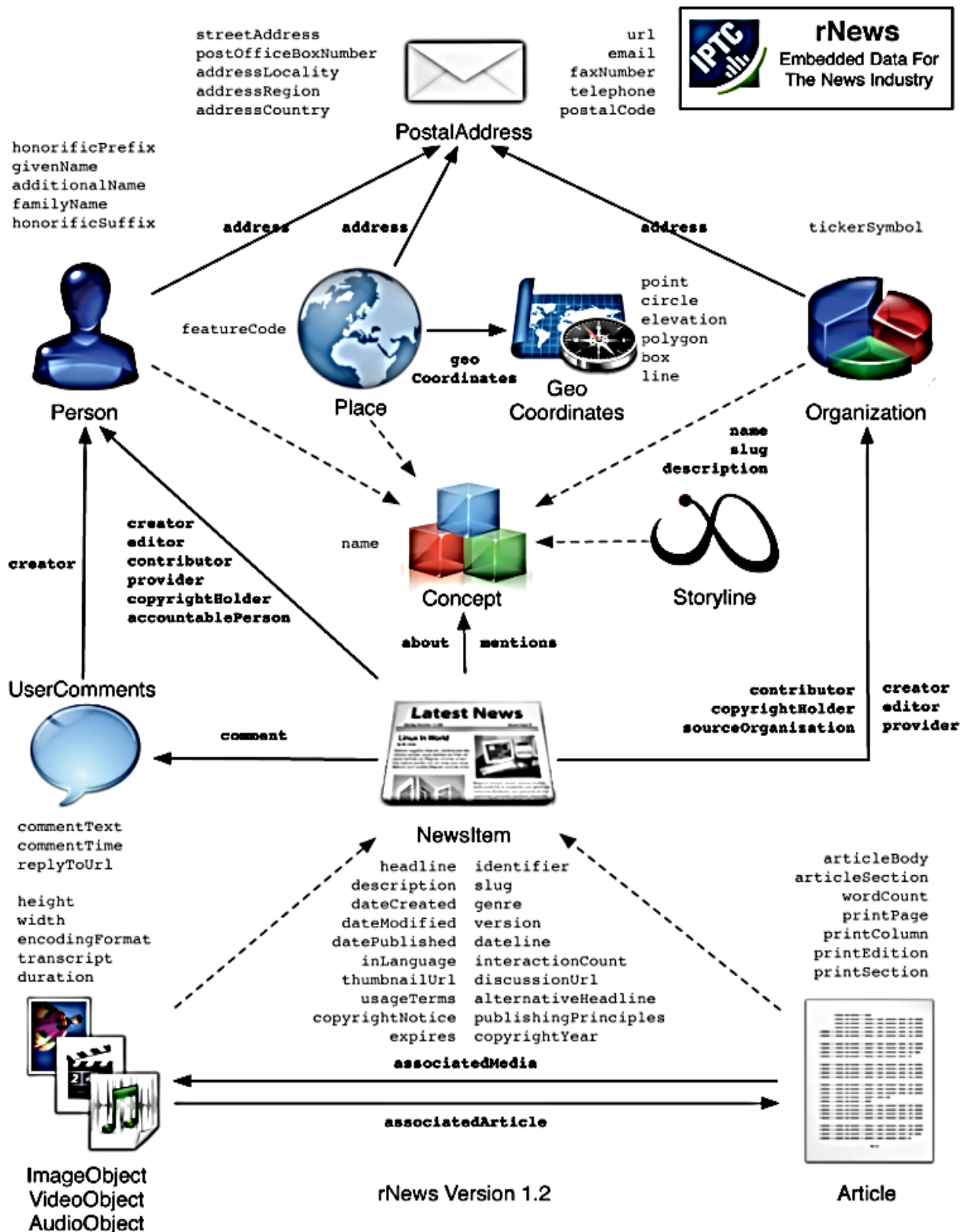


Figure 2. rNews data model (from official IPTC website)

As M. Shilina argues characteristics of all elements of the communication model on the Internet are unique. After all, participants are virtual, information is multimedia, and all communication formats can be implemented. Information revolves as an object of translation, exchange and acquires fundamentally other properties: multidimensionality, multimedia, interactive openness, etc. [5].

N. Akimova notes that indirect communication on the Internet led to the fact that the communicative act shifted accents from the producer to the message [6]. Some arguments to support this idea can be found in this model. After all, to construct an adequate model, in which relations between objects are represented, but the classes do not intersect, two classes are selected as system-building: *NewsItem* one and *Concept* one, which have subclasses and attributes.

Many news sites create their content using subject headings from special controlled vocabularies. Other publishers follow a less formal approach and add annotations with author or editor tags. The *Concepts* class is the basic for denoting all such annotations in the *rNews* data model and it has only one attribute: a name that points to an abstract concept. This class is used to model abstract concepts and is divided into 4 subclasses (*Place*, *Geocoordinates*, *Person*, *Organization* and *Storyline*) which model specific types of real concepts.

Subclass *Person* provides the attributes of exact name specification: the first and second name, surname, honorific prefixes and suffixes (Dr, Mr, PhD, etc.); subclass *Organization* has a single attribute *tickerSymbol* – for exchange traded instruments associated with the organization (by the market identifier code); subclass *Place* is used to annotate geographic objects of the real world; to categorize objects with detailed geographic metadata the subclass *Geocoordinates* is used. These subclasses are associated with subclass *Postal Address* (where street, postal office box number, locality, region, country, as well as telephone, e-mail and URL are given).

Subclass *Storyline* contains three attributes: name, description and slug.

The model provides feedback: the *NewsItem* class is related with *UserComments* class, which has such attributes for modeling comments: the text, the time and the URL. The of authorship combines subclasses *UserComments* and *Person*.

Analysts note that the Internet and social media have significantly reduced the distance between the media and audience. The audience got an opportunity not only to read, but also to evaluate, comment, point out mistakes, to offer new ideas, in the end, to create content [7; 8].

There are some relations between the classes *Concepts* and *NewsItem* and their subclasses – direct and indirect ones. A news item may directly relate to the concept (relation *about*) or contain links or reviews, but not necessarily to the concept (relation *mentions*).

The news item can be produced by a person or organization. Depending on the functions, a person or organization may act as a creator, editor, contributor, provider, copyright holder and accountable person.

The functions of modern media are transformed from elementary informing to interaction, the organization and maintenance process of exchange of information, facts, opinions, emotions and communication by verbal and nonverbal means between communicators.

A typical news site contains news in two main formats – articles and media objects. Articles have attributes such as body (main text), section, word count, print page, column, edition and section. Media objects come in different types: audio, video and images, and consist of attributes such as height, width, encoding format, transcript, and duration. Articles and media objects also have some common attributes.

The *NewsItem* class is divided in four subclasses: *Article*, *Image*, *Video* and *Audio Object*, and combines attributes that are common to both articles and media objects, including the main and alternative headline, text description in free-form, slug, the date of creation, publication and latest modification of the news item, the language, the genre, the thumbnail image associated with the content, copyright notice, usage terms, expires, identifier, version, publishing principles. Very important attributes of *NewsItem* class are discussion and interaction count (e.g., 50 user comments, 20 likes or 30 dislikes) that associated with the *UserComments* subclass.

3. Conclusion

It was found that *rNews* data model is nonlinear, because the communication in new media is positioned as a two-way process of interaction of actors. In the communicative aspect, the model is based on a two-way asymmetric model that fixes the feedback, but retains the leading role of the communicator.

Among the main advantages of the rNews data model users point out: optimizing news search, better analytics and advertising. According to designers «the benefit of getting rNews and hNews adopted by the IPTC is that greater industry support translates into less work for toolmakers: if many news providers support hNews and/or rNews – and do so in very similar ways – then it is easier to build parsers and tools to extract metadata from HTML» [4].

The first one to introduce the rNews standard is The New York Times [8], today it is used by Austrian Presse Agentur, Liberation, Xinhua and others.

The experience of the world's leading agencies can be an important benchmark for implementation the rNews data model in domestic practice and studying the theoretical aspects of its use.

References:

- [1] McCallum K. Future arts, media, and entertainment: seeds for 2020. Retrieved from: <http://www.jfs.tku.edu.tw/17-2/R02.pdf>.
- [2] Степанов В. Н. Этика массовой коммуникации. Средства массовой информации в современном мире: Материалы межвузовской научно-практической конференции. СПб.: Роза мира, 2006, с. 290-296.
- [3] Шилина М. Г. Data journalism – дата-журналистика, журналистика метаданных – в структуре медиакommunikation: к вопросу формирования теоретических исследовательских подходов. Медиаскоп, 2013, № 1. Режим доступа: <http://www.mediascope.ru/node/1263>.
- [4] IPTC Official Website. Retrieved from: <http://www.iptc.org>.
- [5] Шилина М.Г. Интернет-коммуникация и теоретические аспекты исследований масс-медиа. Медиаскоп, 2011, № 4. Режим доступа: <http://www.mediascope.ru/node/972>
- [6] Акімова Н. Психолінгвістичний код сайтів новин: колективне авторство. Психолінгвістика. Психолінгвістика. Psycholinguistics, 2017, №. 22 (1), с. 15-25.
- [7] Суцук М. Як медіа взаємодіють з аудиторією в інтернеті. Режим доступу: <http://watcher.com.ua/2011/05/13/yak-media-vzayemodiyut-z-audytoriyeyu-v-interneti/>.
- [7] Балалаєва О. Ю. Аналіз новинних кодів за стандартами Міжнародної ради з питань преси та телекомунікацій. Науковий вісник Національного університету біоресурсів і природокористування України, 2017, вип. 272, с.103-110.
- [8] rNews is here. And this is what it means. Retrieved from: <https://open.nytimes.com/rnews-is-here-and-this-is-what-it-means-ea25f13417d7>.

Article received: 2018-07-15