



IPTC Mirror

IPTC - INFORMATION TECHNOLOGY FOR NEWS

Content and Coding

The 2003 Autumn Meeting included productive sessions dealing with metadata (including the IPTC Subject Reference System) and with programs for handling various types of structured content - including a new project on Weather Data.

News Metadata

When the IPTC Subject Reference System (SRS) was originally put together it included a number of terms which did not have explanations (which were not specifically required at the time). Working Party Chairman John Minting (UPI) told delegates that a substantial effort to provide appropriate explanations has been made by UPI, with their proposals having been circulated to members for consideration prior to the meeting.

After a brief discussion the proposals were accepted with the exception of the *metal goods and engineering* Subject Matter and Subject Detail entries under the *economy, business and finance* subject. This was because the products concerned are not necessarily made of metal, and a paper will be prepared to deal with this.

Some problems had been encountered in processing SRS explanations that were longer than 255 characters. Where appropriate, revised - shorter - explanations had been proposed (again by UPI) and these were also approved.

Members were informed that a number of additions to the SRS had been made under the "Fast Track" procedure, while some further proposals were under consideration. This process is now complete and all the changes have been taken into a new release of the SRS (Version 11).

Scene TopicSet

A new IPTC TopicSet, initially called ContentCharacteristics, was proposed to provide standard names (with descriptions) for images - for example *headshot, half-length, group, and interior view*. Entries are identified by numeric codes, as with some SRS TopicSets, to allow identification if the names and descriptions are translated into

other languages than the original English.

Provision has been made for extension of the TopicSet to cover other media types, which can be identified by the first two digits of the numeric code. Following discussions the numeric code was extended to allow for the possibility of further details being dealt with. There was general agreement that this TopicSet would prove useful and it was approved under the name "Scene".

A number of additions to the CharacteristicsProperty TopicSet were also considered and approved.

SRS Guidelines

Some detail changes had been made to the draft revision of the Subject Reference System Guidelines to take in points raised at previous Meetings - mainly relating to the procedures for making additions - and these were agreed.

The proposed changes include a section that describes the conditions under which similar terms may appear under different Subjects or Subject Matters. Although the need for this practice was recognised the need to minimise such occurrences was stressed to avoid possible problems when codes are applied.

TopicSets

The relationship between the SRS and other TopicSets was looked at, as there appear to be some inconsistencies in the relationships between them, with the MediaType, NewsItem Type and Genre TopicSets being treated as part of the SRS. It was explained that this was because these sets had been developed when the SRS was primarily intended for use with the IIM (Information Interchange Model), while the other TopicSets had been developed for use with NewsML. The IIM is still a widely used standard so the metadata structure has to remain compatible.

Some changes have been made in conjunction with redevelopment of the IPTC web sites which have a specific area dealing with TopicSets (www.iptc.org/metadata). This includes an overview and brief explanations, along with forms for submitting proposed changes and additions. The information is now held in a database with any modifications being directly reflected in the displayed information. TopicSets are available for download - as NewsML instances - from the IPTC site.

Translations

A Georgian translation of the SRS had been supplied to IPTC by a non-member and while the effort that had been made was appreciated, consideration of this offering raised a number of significant points. For example, input from non-IPTC members has to be sponsored by a member and would need to be properly authenticated before it could be accepted.

Similarly there would also need to be a

Contents

Working Party Reports:

News Metadata	P1
Special Content	P2
EventsML	P2
ProgramGuideML	P2
SportsML	P3
Weather Data	P3
Meetings 2004	P2
Knowledge Management	P6

way of ensuring that the translation is regularly updated in line with developments of the SRS. After some discussion it was decided that the problems would have to be looked at in more detail. One suggestion was that such translations could be made available as unsupported versions, but this would be a significant policy change and so needs proper consideration.

Special Content

EventsML

It was agreed that the EventsML Working Party will be taken into the Special Content Working Party as it will fit in with other work and it should be possible to make better use of available resources. A brief update on progress made with EventsML since the last meeting was provided by Dominic Chan (Canada Newswire).

Initial sets of requirements have been provided by interested parties, consolidated and compared to the existing formats. It appears that vCal and iCal already address 60% to 70% of the IPTC requirements, while both formats have extension facilities. However, neither of these formats is XML based.

Subsequent discussions confirmed that several members - along with outside parties - had a real need for an Events language. The main point to be settled was the best way of doing this. To begin with, would it be better to base the system on an existing system such as vCal (or iCal), or to make a totally fresh start? Then, should the program be a purely internal IPTC development - or would it be better to approach

Provision of Internet connections in the conference room at the Leipzig Meeting was very popular — as can be seen from the cabling! Such connections are now likely to become a regular feature at meetings.



OASIS to see if it would be possible to gain support for a more widely based development project?

It was pointed out that the support of three OASIS members is needed to form a discussion group. Since IPTC and Reuters are members it was decided to try to find an interested third party to form a group, and establish the general level of interest. An additional consideration is that involving OASIS in the development process could help ensure that details of the planned work are more widely known, and reduce the possibility of other initiatives trying to deal with the same requirements.

In parallel with this approach it was decided to start work on the production of a conceptual model for a possible XML Schema. If a co-operative venture is established this model could be made available as a basis for development, otherwise it can form the basis of an IPTC project.

Assuming that a modular design approach is adopted, and sufficient efforts can be made, it is estimated that a completed standard might be available in 12 to 18 months. Since the meeting a new "Conceptual Schema" folder has been added to the files area of the EventsML discussion group (<http://groups.yahoo.com/group/eventml-dev/>). This area is intended to contain prototype XML Schema for EventsML for discussion - it already contains an overview.

ProgramGuideML

Manabu Miyake (The Yomiuri Shimbun), who is the leader of the NSK NewsML Team, provided an update on ProgramGuideML developments. A series of detail changes have been made to the 0.7 version of ProgramGuideML to give a V1.0 draft specification and DTD. This had been circulated prior to the Leipzig meeting and it is hoped V1.0 will be formally approved for release in Spring 2004.

Members were asked to check that the standard conforms to the established IPTC

Meetings 2004

Spring Meeting

Monday 15th to
Thursday 18th March

Athens

This is a four-day meeting with the extra day - the Thursday - having been arranged for discussions with officials about the Olympics.

Annual General Meeting

Tuesday 25th to
Friday 28th May

Hong Kong

The 2nd Chinese NewsML Forum is being held on the Monday before the AGM, and delegates are invited to take part in the Forum.

Autumn Meeting

Wednesday 6th to
Friday 8th October

Amsterdam

format and requirements. In addition the EBU have been asked to check that the technical terms are the most appropriate translations of the original Japanese. It was pointed out that, as with other standards releases, introductory and tutorial documentation would have to be ready for the planned release.

Following an approach from the EBU (European Broadcasting Union) a team representing the ProgramGuideML project visited the EBU in Geneva to discuss the standard and the EBU TV-Anytime project. The TV-Anytime Forum (www.tv-anytime.org) is a consortium of over 150 organisations who are developing open specifications for interoperable and integrated systems that will allow consumer electronics manufacturers, content creators, telecommunications companies, broadcasters and other service providers to make effective use of high-volume digital storage in consumer devices.

The main target is personal video recording systems and an XML based system is being developed for this purpose. It appears that ProgramGuideML and TV-Anytime have different uses, are generally

compatible, and can be used together. Much of the program information used by TV-Anytime is the same as for Program-GuideML and it would be important to have proper mapping between the two metadata sets. With such mapping broadcasting companies that have adopted TV-Anytime could readily deliver ProgramGuideML.

SportsML

It was reported that some minor additions had been made to the sports vocabularies in response to requests from users. One

new area receiving particular attention was horse and harness racing and a subgroup to deal with this had been set up on the yahoo SportsML discussion group. Contact had also been made with companies dealing with racing results.

There had been some discussions with a technical member of the International Olympic Committee (IOC) but progress is likely to be slow as it seems that the IOC have long development times. However, efforts will continue in this area and it was hoped that the Athens meeting might be used to promote SportsML.

Another area receiving attention is fantasy sports, this is seen as having considerable potential as players rely on the available statistics.

There are plans to prepare a SportsML V1.5 proposal for the Athens (Spring 2004) Meeting. To help further development of SportsML there is a need for individual editors who will take responsibility for specific sports plug-ins - there is a possibility that these could be non-IPTC members.

In addition it may be possible to have a draft XML Schema for SportsML available for consideration at the Athens Meeting - in the longer term the move to an XML Schema is seen as having definite advantages as it will allow detailed validation of content.

Alan Karben (XML Team) told members that there are four major sports syndicators in the US market, two of which - AP and XML Team - are already offering SportsML output. Discussions have been held with Sports Ticker (a third syndicator) and there are some hopes that they will start to use the standard. At the moment there does not appear to be any competition to SportsML, although there are some systems for specific sports.

Weather Data

The handling of weather data seems to be gaining in importance for the news industry. One of the driving factors behind this is that newspaper web sites in the USA are looking for material that will attract users, and local weather information is a strong contender. This means that the market would like to see rapid development of an appropriate system.

Some work has already been undertaken with John Iobst (NAA) and Honor Craig-Bennet (PA) having generated a list of requirements and produced a draft DTD which will be circulated for comments. It was decided that this project would also be the first to be carried out using the planned Project Management approach.

Other Content

In the financial area it appears that XBRL (Extensible Business Research Language) is being widely adopted with a lot of financial software products now offering compatibility.

No specific developments in XPRL (extensible Public Relations Language) were reported, but work on the the standard is continuing.

Knowledge Management and Topic Maps

A presentation by Gerhard Köhn, VP Sales, empolis GmbH during the Autumn Meeting provided an overview of Knowledge Management and Topic Maps and the ways they can be used to improve access to information.

As a starting point Mr Köhn explained that although there may be a lot of information sources it can be difficult to find the specific information wanted. In addition there may be a fear that something important may be missed -

this might be because of the quantity of information or because of redundancy in the information with the same concept being identified in different ways. A further problem is that it can be hard to know if an information source can be trusted.

One of the ways of dealing with this is knowledge management which takes both structured and unstructured material and combines it to provide the user with information in a specific context. Core functions in achieving this include the extraction of metadata; summarisation of content, clustering of related content, control of content generation, and identification of areas where there is a gap in the knowledge. This requires efficient searching and a combination of different techniques may be used. In addition collaboration and ideas exchange are key features.

An ontology can be considered as a description of the concepts in a given area of interest, and Topic Maps are an ISO standard for ontology representation. Topic Maps provide information about the occurrences of topics - or the links between topics and relevant resources - and the associations between them. As an example, information on the river Rhone might occur in a map image and there would be an association that the river "flows through" the country France. Another occurrence might be an entry on France in a factbook, with a further association that France and Germany are in the same geographical area. These associations form a complex multidimensional topic space with linking tree structures, and it is then possible to follow the trees backwards and forwards to get an idea of how the information is related.

Mr Köhn concluded with a demonstration of the empolis knowledge suite. This is a XML-enabled platform designed to provide document management, content lifecycle management and information access management. Further information is available from www.empolis.com.

