

*Proseminar WS 04/05:
Auszeichnungssprachen*

XML in Nachrichtenagenturen

von Moritz Hofmann
03.02.2005

Gliederung

- Aufgaben von Nachrichtenagenturen
- Was sind Nachrichten?
- Übersicht wichtiger Nachrichtenagenturen
- Strukturierung von Nachrichten
 - News Industry Text Format (NITF)
 - NewsML

Aufgaben von Nachrichtenagenturen

- Beschäftigen weltweit Journalisten
- Beschaffen und prüfen Nachrichten, auch von externen Quellen & freien Mitarbeitern
- Aufbereitung für Zielgruppe
- Beliefern Zeitungen, Internetanbieter, Fernsehsender damit

Was sind Nachrichten?

- Sollen die Wirklichkeit so gut wie möglich schildern
- Beinhalten aktuelle Informationen, müssen noch am selben Tag weitergeleitet werden
- Können aus Text, Bild, Ton, Video bestehen

Nachrichtenagenturen

- dpa - Deutsche Presse-Agentur
 - Bietet nationale und internationale Nachrichten für alle Ressorts sowie regionale Ergänzungen
 - Seit 1997 dpa-Online für Web-optimierten Content
- Reuters
 - Weltweit größte Agentur, drittgrößte in Deutschland
 - seit 2000 verstärkt online tätig, mittlerweile am meisten gelesene Nachrichtenquelle im Internet

Strukturierung von Nachrichten

- Strukturierung nötig, da Nachrichten auf vielen verschiedenen Medien ausgegeben werden
- Beschleunigt Suche nach Teilen vom Inhalt
- Informationen über Qualitätskontrolle
- schnelle Übermittlung aus Krisengebieten

- IPTC – International Press Telecommunication Council
- internationaler Zusammenschluss von Nachrichtenagenturen seit 1965
- Beschließt Standards zum Nachrichtenaustausch

- NITF – News Industry Text Format
- Wird vom dpa benutzt
- 1996 vom IPTC als Standard entwickelt
- Anfangs in SGML, mittlerweile XML
- An HTML angelehnt – leicht per XSLT konvertierbar, z.B. in HTML, WML, RTF oder andere Ausgabeformate

NITF Aufbau

`<nitf>`

`<head>`

--> Tags für Titel, Datum, Copyright, ...

`</head>`

`<body>`

--> Überschrift, Absätze, Personen, Ort, Adressen, Firmen etc einzeln getagged

`</body>`

`</nitf>`

NITF <head>

```
<?xml version="1.0" ?>
  <!DOCTYPE nitf (View Source for full doctype...)>
  <nitf version="-//IPTC//DTD NITF 3.2//EN" change.date="October 10, 2003" change.time="19:30">
  <head>
  <title>Norfolk Weather and Tide Updates</title>
  <tobject tobject.type="news">
    <tobject.property tobject.property.type="feature" />
    <tobject.subject tobject.subject.refnum="17000000" tobject.subject.type="Weather"
  tobject.subject.ipr="IPTC" />
    <tobject.subject tobject.subject.refnum="17004000" tobject.subject.matter="Statistics"
  tobject.subject.ipr="IPTC" />
    <tobject.subject tobject.subject.refnum="04001002" tobject.subject.detail="Fishing Industry"
  tobject.subject.ipr="IPTC" />
  </tobject>
  <evloc city="Norfolk" state-prov="VA" iso-cc="US" />
  <del-list>
    <from-src src-name="iptc.org" level-number="1" />
    <from-src src-name="xmlteam.com" level-number="2" />
  </del-list>
```

...

NITF <body>

```
...
<body>
  <body.head>
    <headline>
      <hl1>Weather and Tide Updates for Norfolk</hl1>
      <hl2>A sample, fictitious NITF article</hl2>
    </headline>
    <note noteclass="editorsnote" type="npa">
      <body.content>
        <p>This sample article was created completely from scratch in order to illustrate various
          features of NITF. Parts of it are somewhat contrived, in order to illustrate as much of
          the DTD as possible.</p>
      </body.content>
    </note>
  ...
```

Weather and Tide Updates for Norfolk

A sample, fictitious NITF article

Editor's Note: This sample article was created completely from scratch in order to illustrate various features of NITF. Parts of it are somewhat contrived, in order to illustrate as much of the DTD as possible.

By Alan Karben
NITF Network News Online

[ScreamingMedia, Inc.](#) Newport News, VA

Superb weather in Norfolk today. Tides coming in and out as per usual.

The weather was superb today in Norfolk, Virginia. Made me want to take out my boat, manufactured by the [Acme Boat Company](#).

Tides in Norfolk are running normal today. This weeks article highlights many of this week's fishing issues, and



- Von Reuters entwickelt
- 2000 vom IPTC als Standard verabschiedet
- Zum strukturieren von Multimedia-News
- Ziele: medienunabhängig, automatische Verarbeitung erleichtern, vielfältige Ausgabemedien
- Vollständig in XML umgesetzt

NewsML Struktur

- NewsML
 - Catalog (z.B. URL)
 - NewsEnvelope (Nachrichtenagentur, Datum, Empfänger)
 - TopicSet (Titel, Beschreibung)
 - NewsItem (Inhalt)

NewsML Beispiel

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<NewsML>
  <Catalog Href="http://www.afp.com/dtd/AFPcatalog.xml" />

  <NewsEnvelope>
    <DateAndTime>20030319T000031Z</DateAndTime>
    <NewsService FormalName="FRSGL" />
    <Priority FormalName="3" />
  </NewsEnvelope>

  <NewsItem>
    <Identification>
      <NewsIdentifier>
        <ProviderId>afp.com</ProviderId>
        <DateId>20030319</DateId>
      </NewsIdentifier>
      <NameLabel>Irak - UNO</NameLabel>
    </Identification>

    <NewsComponent>
      <NewsLines>
        <HeadLine>Bagdad lehnt Ultimatum ab</HeadLine>
        <DateLine>BAGDAD, 18 März (AFP) -</DateLine>
        <CopyrightLine>© 2002 AFP</CopyrightLine>
      </NewsLines>
    </NewsComponent>
  </NewsItem>
</NewsML>
```

Literaturverzeichnis

- Höhne, H: Report über Nachrichtenagenturen, 1984
- Ministerium für Kultur Baden Württemberg:
www.mediaculture-online.de
- dpa Website: www.dpa.de/de/unternehmenswelt
- Reuters Website: about.reuters.com
- Rall, P: Die Position von Reuters im Nachrichtenmarkt, 2003
- International Press Telecommunication Council: www.iptc.org
- NITF: www.nitf.org
- Reuters NewsML: about.reuters.com/newsml
- NewsML: www.newsml.org