

Considerations for the Semantic Markup with the NITF

Version: 1 February, 1998

INTRODUCTION

When coding semantic units of a text using the NITF Document Type Definition and the content elements defined there, a number of questions need to be answered if we want to arrive at a consistent list of items that may be turned into a useful index for later retrieval. Speaking in general terms, there are three main areas of concern: the correct and consistent classification of semantic units, the proper degree of specificity, differentiation and granularity, and the relationships that exist between content elements that stand in close vicinity within the text. These issues do not allow solutions once and for all, but need to be discussed in connection with the specific semantic categories.

A question that may be addressed and decided before we start discussing the problems connected with the different semantic categories concerns the selection of term boundaries. This is a more formal matter that may be handled by a system:

As a principle, blanks are taken to delimit the boundaries of words. This means, inflected word forms (e.g. *London's*, *ministers*) and hyphenated phrases (*London-based*, *Simpson-case*) are not to be interspersed with tags in order to cut off the normalised form of an expression. Instead, we suggest using a full-form dictionary and to introduce an attribute storing the normalised form to solve problems which may arise from this procedure. Leading and trailing blanks are associated with the text, not the tags.

The issues that we need to tackle if we want to achieve consistency in tagging are, in particular: the semantic content of each tag set (tag content); how to deal with semantic ambiguity of expressions and thus with the possibility of assigning a semantic unit to more than one tag set (multiple categorisation), the intended degree of specificity in coding phrases (specificity), the interrelation of content elements that are associated with each other by closeness within a specific context (serial relations), and how to deal with the different forms of contracted word groups (contracted word groups). What these issues involve will be discussed next.

Tag content

Tag content refers to the character data that are delimited from the surrounding text by begin and end tags of a given content element. Determining the tag content answers the question as to what exactly we wish to group under a given semantic category.

As to a correct classification, i.e. which tag to choose for a given content, there may be different concerns:

At times, it is the very object that seems to forbid a clear assignment to one of the given categories. Possibly, because its qualities are not well defined and it may well go into either of two related categories (e.g. STATE and REGION), or it shifts roles with the context (an airport may figure as a LOCATION in one story and as an ORGANISATION in another). The object might also allow for multiple categorisation (*Sharon* may be a personal name or the name of a river), or the category is so broad that the tag content seems awkward or wrong (a village within the CITY tag).

It may further be argued, that semantic tagging, in order to support later retrieval, should, probably, go beyond tagging proper names and comprise coding object types, like types of firms (*insurance company*, *diversified utility group*), types of locations (*German resort*, *tent city*), types of events (*trial*,

fire, aeroplane crash), and types of objects (*newspaper, musical*). This, however, would make a list of items grouped under one category rather broad in spectrum.

All these questions will eventually have to be answered conclusively, if a consistent tagging is to be achieved.

Multiple categorisation

Quite frequently, a semantic unit may be assigned to more than just one category. We address this as multiple categorisation. Actually, it is a specific problem concerning the tag content, but as its discussion may require quite a lot of space with some categories, we treat it under a separate heading. Multiple categorisation may occur for different reasons:

With homographic terms (same spelling, different meaning), it is by pure coincidence that the same word or phrase may refer to different concepts or objects. *Washington*, for instance, may be marked up as a city, a mountain, a given name, or a family name.

With the use of figurative language, in principle, the term for something familiar gets substituted by something less familiar (*Bremen* instead of *Sportverein Werder Bremen*), something concrete by something abstract (*White House* for the *American administration*), the whole by a part (*the Defence Department said ...* instead of the *spokesman of the Defence Department said*).

It also occurs that objects shift their roles depending on the context. This, for example, happens with countries or other political entities, which mainly are referred to as locations, but may also function as organisations, if, for instance, they are mentioned as member states of a larger organisation.

A decision how to deal with the possible multiple categorisation of a given semantic unit can only be made considering its context, and rules may be necessary to achieve consistent coding.

Specificity

A base concept (head or core of a phrase) may be extended by pre modifications (extensions to the left) and post modifications (extensions to the right). Each added modification increases the specificity of the concept. Thus, a base concept, like the *president*, may be extended to the *former president*, the *indicted president*, the *president-designate*, the *President of the United States of America*, the *former President of the United States of America*. It can be argued that, in order to enhance precision, the lowest possible level of abstraction should be selected, i.e. the tags should capture the most specific expression occurring in the text.

However, the modifiers are of a very different nature. They may be necessary for an unambiguous reference or they may convey redundant information put in as an extra for the reader's information. In the above example, *former* and *United States of America* may be seen as classifying modifiers (the *former president* is no longer a *president*. The *President of the United States* needs to be distinguished from *presidents* of other countries), whereas *indicted* is a qualifying attribute, not an integral part of the concept, and merely added for the reader's orientation. The problem of qualifying or classifying modifiers is particularly relevant to locations, which are heavily used as modifiers in news stories. They may either precede the headword (*the Atlanta Olympic Games; Nürnberger Bundesanstalt für Arbeit*) or follow it (*the Olympic Games in Atlanta; Bundesanstalt für Arbeit in Nürnberg*).

Serial relations

If the issue of specificity addresses what to capture with a single tag and what to leave to another tag, serial relations address the questions of how relationships between different types of semantic units that occur together in close textual vicinity may be made explicit by tagging. SGML, originally a document description language representing a document's structure as a whole-part hierarchy, has no notion of semantic relations in the sense addressed here. Thus, substructures associating the

specified classes of semantic units in NITF remained rather coarse. An example for this is the modelling of the LOCATION element, with its inner elements SUBLOCATION, CITY, REGION, STATE, COUNTRY. These can be taken to represent an order of increasing size and may allow us to interpret a city and country within the LOCATION to read as: a certain city is located within a certain country. To preserve this information would go beyond building an index (with the implied simple message that city and country both are locations) and to store (still simple) facts, such as: a certain city is part of a certain country. A similar example is to use the FUNCTION tag outside and inside the PERSON tag. When inside the PERSON tag, the FUNCTION is associated with the personal name within the tag, which means: this specific person performs this specific function. The problem connected with capturing or not capturing modifiers within a tag, could - may be - lessened, if we had another look at the question of what are the most frequent modifying semantic units and which other semantic units do they modify. LOCATION, in running text, for example, seems to be more often used as a modifier than in its own right, as in, e.g. the *Atlanta Olympic Games*, the *Olympic Games in Atlanta*.

Contracted word groups

Co-ordinated word groups are often contracted to avoid repetition. Thus, an ellipsis occurs, as in *the ministers Kinkel and Seehofer*, and causes a markup problem because of the omitted word or phrase. It might be tackled by introducing an attribute that stores the left-out expression to allow an expansion into the complete form: *minister Kinkel and [minister] Seehofer*.

Semantic units defined in the NITF

The semantic units or content elements specified in NITF are:

PERSON: personal names

FUNCTION: a person's position or role

ORG: an organisation name

LOCATION: a place name

EVENT: a news-relevant event

OBJECT.TITLE: the name of a news-relevant object

CHRON: a time expression

MONEY: a monetary item

NUM: a numeral expression

Q, BQ: quoted information

These semantic units will be discussed in this paper in the above order. The combination or serial relation between PERSON and FUNCTION, because of its importance, is given a special section following the individual discussion of both elements.

The different content elements will be discussed in terms of the above-addressed issues and usage examples will be given to illustrate these. The examples consist of phrases taken from English and German news messages. For better transparency, we have limited tagging to the text that is relevant to the points we make.

Where appropriate, some suggestions are made as to possible extensions to the NITF standard.

PERSON <PERSON> ... </PERSON>

The PERSON tag may occur anywhere in running text.

Tag content

To tag a person typically means tagging a name that refers to this specific person. As we do not intend with these guidelines to elaborate fully on the difficult topic of name indexing, only a few, more obvious observations are made.

The PERSON tag allows to either include mere character string and/or to differentiate a personal name into forenames and surnames. NITF uses the terms: given name (NAME.GIVEN) and family name (NAME.FAMILY). The family name, however, is not necessarily identical with an inherited name.

Prefixes to personal names, as in *Mr. Blair*, *Frau Merkel*, and suffixes, as in *John F. Kennedy Jr.*, *Johannes Paul II* are not explicitly tagged as such. They remain within the PERSON tag.

Furthermore, a person's name when mentioned in a news message is often prefaced or followed by a function associated with this person (*Whitewater special prosecutor Kenneth Star*, *Stephen Dorrell, the Health Secretary*). NITF allows making this relation explicit by putting the FUNCTION tag inside the PERSON tag.

Existing, historical and fictitious persons should be tagged. Non-real persons should possibly be marked as such. Some personalities, of course, may belong to both the real and the non-real realm (*Kaspar Hauser*, *Mary Stuart*).

Suggested modification to the element: Introduce attribute named "status" with living, fictitious, and historical as its value, with default value: living.

GIVEN NAMES + FAMILY NAMES

<PERSON><NAME.GIVEN></><NAME.FAMILY></></PERSON>

A standard form to refer to persons is by citing their given and family names.

<PERSON><NAME.GIVEN>Tony</NAME.GIVEN><NAME.FAMILY>Blair</NAME.FAMILY></PERSON>

<PERSON><NAME.GIVEN>Angela</NAME.GIVEN><NAME.FAMILY>Merkel</NAME.FAMILY></PERSON>

If there are two or more given names, no distinction is made in the NITF as to middle names. The forenames are considered a sequence of given names, including initials.

<PERSON><NAME.GIVEN>John</NAME.GIVEN><NAME.GIVEN>Fitzgerald</NAME.GIVEN><NAME.FAMILY>Kennedy</NAME.FAMILY></PERSON>

<PERSON><NAME.GIVEN>John</NAME.GIVEN><NAME.GIVEN>F.</NAME.GIVEN><NAME.FAMILY>Kennedy</NAME.FAMILY></PERSON>

Some persons are referred to by using two surnames, e.g., their birth name and their married name, they are considered a sequence of family names:

<PERSON><NAME.GIVEN>Debbie</NAME.GIVEN><NAME.FAMILY>Rowe</NAME.FAMILY>

<NAME.FAMILY>Jackson</NAME.FAMILY></PERSON>

FAMILY NAME <PERSON><NAME.FAMILY></NAME.FAMILY></PERSON>

Frequently, in particular, when a person's full name has already been given, only the surname is mentioned.

<PERSON><NAME.FAMILY>Jackson</NAME.FAMILY></PERSON>

<PERSON><NAME.FAMILY>Waigel</NAME.FAMILY></PERSON>

Family names may come as combinations of surnames, either with or without a hyphen:

<PERSON><NAME.FAMILY>Warwick-Thompson</NAME.FAMILY></PERSON>

<PERSON><NAME.FAMILY>Hamm-Brucher</NAME.FAMILY></PERSON>

Family names may also include particles:

<PERSON><NAME.FAMILY>von Hanstein</NAME.FAMILY></PERSON>

<PERSON><NAME.FAMILY>de Klerk</NAME.FAMILY></PERSON>

There are complicated name forms or unfamiliar names (e.g., Arabic, Japanese) that require special knowledge for a correct coding.

<PERSON><NAME.FAMILY>Huschke von Hanstein</NAME.FAMILY></PERSON>

<PERSON><NAME.FAMILY>Ng</NAME.FAMILY></PERSON> (from Chinese name: Ng Lap Seng)

Plural forms of family names exist, when two or more members of the whole family are collectively addressed by using the plural form of a family name. According to the rule that tags are not to disrupt word boundaries the plural form as such is tagged as the family name.

the <PERSON><NAME.FAMILY>Blairs</NAME.FAMILY></PERSON>

Die beiden<PERSON><NAME.FAMILY>Vogels</NAME.FAMILY></PERSON>

GIVEN NAME <PERSON><NAME.GIVEN></NAME.GIVEN></PERSON>

At times, only the given name is cited and has to be tagged accordingly.

<PERSON><NAME.GIVEN>Elvis</NAME.GIVEN></PERSON>

<PERSON><NAME.GIVEN>Steffi</NAME.GIVEN></PERSON>

Name prefixes and suffixes: Name prefixes include polite addresses (*Mr, Mr., Frau, etc.*), academic titles (*Dr., Prof. etc.*), titles of British royalty and nobility (*Queen, Dame, Sir, Lord, Lady, etc.*), they are included in the PERSON tag as PCDATA without further identification.

<PERSON>Lord <NAME.FAMILY>Bingham </NAME.FAMILY>of Cornhill </PERSON>

<PERSON>Queen <NAME.GIVEN>Elisabeth </NAME.GIVEN></PERSON>

<PERSON>Frau<NAME.FAMILY>Merkel </NAME.FAMILY></PERSON>

<PERSON>Frau Dr. <NAME.FAMILY>Merkel </NAME.FAMILY></PERSON>

Name suffixes include abbreviations for degrees (*Ph.D., M.D.*), numbers (*as in Elisabeth II*) or epithets like *Jr. Sr.* They are also included within the PERSON tag, but not tagged explicitly.

<PERSON>Queen <NAME.GIVEN>Elisabeth</NAME.GIVEN> II</PERSON>

der Verleger <PERSON><NAME.FAMILY>Unselde</NAME.FAMILY> Junior</PERSON>

Contracted word groups

If two persons with the same family name are mentioned, they may be tagged like this:

<PERSON><NAME.GIVEN>Bill</NAME.GIVEN></PERSON> and

<PERSON><NAME.GIVEN>Hillary</NAME.GIVEN>

<NAME.FAMILY>Clinton</NAME.FAMILY></PERSON>

Multiple categorisation

person | object: Possible multiple categorisation occurs when personal names are used figuratively in sentences like *Gunman takes £ 700,000 Picasso*. In this case, the personal name is a convenient short form for *a work of Picasso's* and should be given the OBJECT.TITLE tag.

<PERSON><NAME.FAMILY>Picasso</NAME.FAMILY></PERSON> painted the picture

Gunman takes £ 700,000 <OBJECT.TITLE>Picasso</OBJECT.TITLE>

<PERSON><NAME.FAMILY>Picasso</NAME.FAMILY></PERSON> malte das Werk ...

£ 700.000 teurer <OBJECT.TITLE>Picasso</ OBJECT.TITLE> gestohlen

person | organisation: Organisations that carry the family name of their founders should be tagged as organisations:

the powerful fund managers and unit trust providers such as <ORG>Schroders</ORG>

<PERSON><NAME.FAMILY>Langenscheidt </NAME.FAMILY></PERSON> sprach auf der Buchmesse

<ORG>Langenscheidt</ORG> hat weltweit 1 300 Mitarbeiter.

Serial relations

person & organisation: A person's affiliation is often mentioned in close textual vicinity of the person.

<PERSON><NAME.GIVEN>Andrew</NAME.GIVEN> <NAME.FAMILY>Warwick-Thompson
</NAME.FAMILY></PERSON>, of <ORG>Bacon & Woodrow</ORG>, who ...

<PERSON><NAME.GIVEN>Michael</NAME.GIVEN>
<NAME.FAMILY>Hughes</NAME.FAMILY></PERSON>, of <ORG>West Ham</ORG>, is ...

<PERSON><NAME.GIVEN>Klaus</NAME.GIVEN>
<NAME.FAMILY>Sprick</NAME.FAMILY></PERSON> von der <ORG>dpa</ORG>

The connection between FUNCTION and PERSON tags, which we consider a serial relation, will be discussed in an extra section below (see PERSON & FUNCTION).

Notes: Personal names used as modifiers of nouns (*Simpson case*, *Barschel-Affäre*) and thus become part of a compound noun group should not be tagged separately as persons. In the following two examples, the whole phrase should be tagged as EVENT.

<EVENT>Simpson case</EVENT>

<EVENT>Barschel-Affäre</EVENT>

Persons are not always addressed by their personal name. Quite frequently, once a person is introduced by function and name, the function is used synonymously to refer to this person (*Prime Minister John Major -> the Prime Minister*). The FUNCTION tag should then be used.

The <FUNCTION>Prime Minister</FUNCTION> then went on

Der <FUNCTION>Ministerpräsident</FUNCTION> beschloß

Some persons remain unnamed, because they are only identified via their relation to better known persons (*Michael Jackson's wife*, *Deckerts Anwalt*) or by a paraphrase that is chosen for the context. As this is a reference to a specific individual with a name of his/her own, the actually occurring name of the better known person is not marked as such, but the whole expression is coded with the FUNCTION tag.

<FUNCTION>Michael Jackson's wife</FUNCTION>, <FUNCTION>Jackson wife</FUNCTION> (from headline)

A <FUNCTION>Roman Catholic archbishop</FUNCTION> who wants to tear down most of a church building that is protected by a historic structure

<FUNCTION>Simpsons Verteidiger</FUNCTION>

To identify a person by his/her forename in combination with his/her relationship to a more familiar person may be considered as partial anonymity. We suggest tagging the relationship to the familiar person with the FUNCTION tag.

<PERSON><FUNCTION>Jackson's wife</FUNCTION>
<NAME.GIVEN>Debbie</NAME.GIVEN></PERSON>

The problem here is that relationships - if they are FUNCTIONS - are of a completely different type as e.g. political or professional functions.

FUNCTION <FUNCTION></FUNCTION>

The FUNCTION tag may occur anywhere in the running text or inside the PERSON tag.

Tag content

Typically, the FUNCTION tag marks full official titles, such as the correct denotation of political, commercial, clerical, military, civil appointments: *e.g. Social Security Secretary, Lord Chief Justice, executive director; Regierungssprecher, Präsident des Bundesverbandes der Textilindustrie, SPD-Fraktionsvorsitzender.*

<FUNCTION>Agriculture Minister</FUNCTION>
<FUNCTION>chief executive of TrizecHahn</FUNCTION>
<FUNCTION>Detective Superintendent</FUNCTION>
<FUNCTION>Regierungssprecher</FUNCTION>
<FUNCTION>Präsident des Bundesverbandes der Textilindustrie</FUNCTION>

It also codes the synonyms and journalistic variants used for official full titles, including their broader terms (*President* instead of *the President of the United States of America*, *Präsident* instead of *Präsident des Bundesverbandes der Textilindustrie*) and titles in the typical news paper jargon, such as the "*rogue trader*" *Nick Leeson*, "*King of Pop*" *Michael Jackson*; *Vize, CDU-Chef, Einheitskanzler.*

the long investigation of the <FUNCTION>President's</FUNCTION> real estate transactions

ausgerechnet der SPD-Abgeordnete Klaus Hahnzog hatte sich geweigert, die Aufhebung der Immunität des <FUNCTION>CSU-Kollegen</FUNCTION> auf die Tagesordnung zu setzen

The FUNCTION tag is also used to identify members of a profession (job titles): *lawyer, journalist, detective, coach, actor; Rechtsanwalt, Journalist, Detektiv, Trainer, Schauspieler, Schauspielerin.*

an <FUNCTION>auto mechanic</FUNCTION> identified McVeigh as the person who allegedly

rented the Ryder truck used in the deadly explosion

pfllegt der meist braungebrannte <FUNCTION>Rechtsanwalt</FUNCTION> seine Vorlieben

Close family relations and relations, such as *father, wife; Vater, Ehefrau*, or otherwise important relations, such as *consultant, employer; Berater, Auftraggeber*, are also tagged as FUNCTION.

his <FUNCTION>wife</FUNCTION>, his <FUNCTION>lawyer</FUNCTION>

seine <FUNCTION>Ehefrau</FUNCTION>, ihr <FUNCTION>Anwalt</FUNCTION>

German and English sports news typically contain - as a stylistic feature - geographical roles as synonymous expressions for personal names, such as *Bremer* for *Marko Bode* or *Hinterzartener* for *Dieter Thoma*. These geographical roles could also be tagged as FUNCTION.

The <FUNCTION>Scot</FUNCTION> only hit the post.

Der <FUNCTION>Bremer</FUNCTION> traf nur den Pfosten.

The FUNCTION tag may further be used to identify important (named) or indicative (unnamed) players in recurring news-relevant scenarios, such as, elections, trials, accidents and natural catastrophes, business, cultural or sport events. In these instances, the FUNCTION is very close to a role; i.e. the same person may change his/her function dependent of the context. In law suits, for examples, these types of players are: *defendant*, *Chief Justice*, *prosecutor*, *witness*, *juror*; *Angeklagter*, *Vorsitzender Richter*, *Kläger*, *Zeuge*, *Schöffe*.

<PERSON><FUNCTION>special prosecutor</FUNCTION> <NAME.GIVEN>Kenneth</NAME.GIVEN>
<NAME.FAMILY>Starr</NAME.FAMILY></PERSON>

Istanbul's Besiktas will punish the <PERSON><FUNCTION>striker</FUNCTION>
<NAME.GIVEN>Daniel</NAME.GIVEN> <NAME.FAMILY>Amokachi</NAME.FAMILY></PERSON> after a trip home to visit his sick mother almost made him miss a key match.

Note: To avoid cluttering the messages with non-relevant functions, fixed lists of functions that relate to the different scenarios may be created. These would exclude less significant players, such as spectators; Anwesende, Gaffer, Schaulustige, Publikum etc. The lists may include, for example, the following players:

elections: candidate, winner, loser, Social Democrat; Kandidat, Gewinner, Verlierer, Sozialdemokrat

law suits: defendant, Chief Justice, witness, juror; Angeklagter, Vorsitzender Richter, Kläger, Rechtsbeistand, Beisitzer, Zeuge, Schöffe

accidents, catastrophes, crimes: dead, victim, hostage, police, criminal; Tote(r), Opfer, Geisel, Geiselnehmer, Entführer, Polizist, Verbrecher

sport events: winner; Sieger, Weltmeister, Europameister, Titelverteidiger, Eisprinzessin

Contracted word groups

Several functions may be listed and contracted into one phrase. Depending on the type of contraction, the tagging must vary: within the context of German politics the *Frauen- und Jugendministerin* refers to one and the same person (Claudia Nolte), the *Finanz- und Wirtschaftsminister* (Waigel, Rexrodt) to two different persons. The tagging in both cases must look like this:

<FUNCTION>Foreign and Home Secretaries</FUNCTION>

Die <FUNCTION>Frauen- und Jugendministerin</FUNCTION> hielt sich bedeckt

<FUNCTION>Finanz- und Wirtschaftsminister</FUNCTION>, waren schon abgereist

Suggested modification to the element: Introduce an attribute with the name FULL. The value must be inserted in free text.

<FUNCTION FULL="Finanzminister">Finanz-
</FUNCTION>und<FUNCTION>Wirtschaftsminister</FUNCTION>

Multiple categorisation

function | family name: Because so many family names were once derived from a person's activity or profession, the possible multiple categorisation of professional titles must be dealt with in German-language news where both names and nouns are capitalised.

<FUNCTION>TV-Ganovenjäger</FUNCTION>
<PERSON><NAME.FAMILY>Zimmermann</NAME.FAMILY></PERSON> geht in Pension.

In principle, each of the nouns belonging to the German every-day vocabulary may also be family name.

Specificity

The headword or nucleus of a function may be preceded or succeeded by modifiers that add more specific information. Pre modification of functions basically consists of an adjective that qualifies the function or a noun, mostly an organisation or location. Post modification consists of a noun or a noun group, mainly organisations and locations. A choice needs to be made as to which modifier must be included in the tag. We agreed on the following rules:

Adjectives premodifying FUNCTION: Adjectives preceding a function are included in the tag

1. If they are part of the official title:

<FUNCTION>Prime Minister</FUNCTION>, <FUNCTION>Chief Justice</FUNCTION>
<FUNCTION>Erster Bürgermeister</FUNCTION>, <FUNCTION>Regierender
Bürgermeister</FUNCTION>, <FUNCTION>Stellvertretender Vorsitzender, <FUNCTION>Vorsitzender
Richter</FUNCTION>

2. If they classify the function unambiguously (classifying adjective):

<FUNCTION>former World Cup captain</FUNCTION> Lothar Matthaeus
<FUNCTION>ehemaliger Bundeskanzler</FUNCTION>, <FUNCTION>designierter
Bundespräsident</FUNCTION>, <FUNCTION>stellvertretender Bürgermeister</FUNCTION>

However, adjectives that add information to a function beyond the identifying reference are not tagged (qualitative adjectives):

the new <FUNCTION>Prime Minister</FUNCTION>
amtierender <FUNCTION>Oberbürgermeister</FUNCTION>, neuer
<FUNCTION>Oberbürgermeister</FUNCTION>, dienstältester <FUNCTION>Minister</FUNCTION>

3. If they indicate nationality or origin (geographical adjectives):

<FUNCTION>Scottish Secretary</FUNCTION>;
<FUNCTION>luxemburgischer Ministerpräsident</FUNCTION>, <FUNCTION>Frankfurter
Oberbürgermeister</FUNCTION>

Particles and adjectives that have become part of the function term are included in the tag.

<FUNCTION>Ex-minister</FUNCTION>, <FUNCTION>then-President</FUNCTION>,
<FUNCTION>president-designate</FUNCTION>
<FUNCTION>Ex-Regierungschef</FUNCTION>, <FUNCTION>Alt Bundeskanzler</FUNCTION>

Organisations premodifying FUNCTION: Organisations prefacing a function are included in the tag. In German, they are connected to the headword by a hyphen.

John Monks, <FUNCTION>TUC General Secretary</FUNCTION>, said
<FUNCTION>FDP-Vorsitzende</FUNCTION>, <FUNCTION>Frontal-Moderator</FUNCTION>

Personal names premodifying FUNCTION: Proper names preceding a function are coded within FUNCTION, if the whole expression refers to a further unnamed person [cf. PERSON]:

<FUNCTION>Mielkes Anwalt</FUNCTION>

This rule also comprises family relations that are coded as functions:

<FUNCTION>Deckerts Bruder</FUNCTION>

Organisations postmodifying FUNCTION: Quite frequently, organisations are used as post modifiers of FUNCTION:

<FUNCTION> Speaker of the British Parliament</FUNCTION>

<FUNCTION>Bundesvorsitzender der SPD</FUNCTION>

<FUNCTION>Richter am Bundesverfassungsgericht</FUNCTION>

<FUNCTION>Sprecher der Leitstelle der Bezirksregierung</FUNCTION>

Locations postmodifying FUNCTION: Locations also occur as post modifiers to functions. There is, however, a difference between the *Oberbürgermeister von Frankfurt* and the *Bundeskanzler aus Oggersheim*. Although both are locations, the former adds information necessary for identification, the latter carries additional information. This requires a different markup.

the <FUNCTION>Mayor of Lincoln</FUNCTION>

der <FUNCTION>Oberbürgermeister von Frankfurt</FUNCTION>

der <FUNCTION>Bundeskanzler</FUNCTION> aus
<LOCATION><CITY>Oggersheim</CITY></LOCATION>

PERSON & FUNCTION <PERSON><FUNCTION></FUNCTION></PERSON>

Quite frequently, a person's name and his/her function occur closely connected in the text. If a personal name is either directly prefaced (pre modification) or followed (post modification) by the associated function, the FUNCTION tag is within the PERSON tag.

Specificity

FUNCTION premodifying a personal name: A function used as a premodifier of a personal name is enclosed within the PERSON tag.

<PERSON><FUNCTION>Labour Prime Minister</FUNCTION> <NAME.GIVEN>Clement</NAME.GIVEN>
<NAME.FAMILY>Attlee</NAME.FAMILY></PERSON>

<PERSON><FUNCTION>vice-chairman</FUNCTION> <NAME.GIVEN>Alan</NAME.GIVEN>
<NAME.FAMILY>Wadey</NAME.FAMILY></PERSON>

<PERSON><FUNCTION>Bundeskanzler</FUNCTION> <NAME.GIVEN>Helmut</NAME.GIVEN>
<NAME.FAMILY>Kohl</NAME.FAMILY></PERSON>

<PERSON><FUNCTION>Regisseur</FUNCTION> <NAME.GIVEN>Peter</NAME.GIVEN>
<NAME.FAMILY>Palitzsch</NAME.FAMILY></PERSON>

FUNCTION postmodifying a personal name: If the person's name is introduced first and the function is added as a post modifier, it is also included in the PERSON tag.

<PERSON><NAME.GIVEN>Michael</NAME.GIVEN> <NAME.FAMILY>Heseltine</NAME.FAMILY>, the
<FUNCTION>Deputy Prime Minister</FUNCTION></PERSON>, welcomed

<PERSON><NAME.GIVEN>Helmut</NAME.GIVEN> <NAME.FAMILY>Kohl</NAME.FAMILY>, der
langj%hrige <FUNCTION>Bundeskanzler</FUNCTION></PERSON>

FUNCTION pre- and postmodifying personal names: A case of very condensed information is the following example, where the FUNCTION is used for pre- and post- modification of a personal name. We suggest tagging it like this:

[He] would be a high-profile casualty, along with the <PERSON> <FUNCTION>Liberal
Democrat</FUNCTION><FUNCTION>Earl</FUNCTION> <NAME.FAMILY>Russell</NAME.FAMILY>,
<FUNCTION>son of Bertrand Russell</FUNCTION>, and the <FUNCTION>Earl of Caernarvon

</FUNCTION>, a <FUNCTION>crossbencher</FUNCTION>, who is <FUNCTION>racing manager to the Queen</FUNCTION></PERSON>.

FUNCTION premodifying given name: In German, *Frau* is not only used as a polite address, it is also a short form for *Ehefrau* and thus denotes a family relation. Phrases like *seine Frau Margot* may be tagged like this:

his <PERSON><FUNCTION>wife </FUNCTION> <NAME.GIVEN>Hillary</NAME.GIVEN></PERSON>
seine <PERSON><FUNCTION>Frau </FUNCTION> <NAME.GIVEN>Margot
</NAME.GIVEN></PERSON>

In special cases, a title may be followed by a first name only, such as, *Queen Elisabeth, Mutter Theresa*. We suggest coding this as in the preceding example:

<PERSON><FUNCTION>Prince</FUNCTION> <NAME.GIVEN>William</NAME.GIVEN></PERSON>
<PERSON><FUNCTION>Mutter</FUNCTION> <NAME.GIVEN>Theresa</NAME.GIVEN></PERSON>
<PERSON><FUNCTION>K^{nigin}</FUNCTION> <NAME.GIVEN>Elisabeth</NAME.GIVEN></PERSON>

Personal Name postmodifying FUNCTION: If the person is used as an apposition to the function, separated by a comma, the connection is considered close enough, to tag person and function together.

the <PERSON><FUNCTION>Prime Minister</FUNCTION>,<NAME.GIVEN>Tony</NAME.GIVEN>
<NAME.FAMILY>Blair</NAME.FAMILY></PERSON>
der langjährige
<PERSON><FUNCTION>Bundeskanzler</FUNCTION>,<NAME.GIVEN>Helmut</NAME.GIVEN>
<NAME.FAMILY>Kohl</NAME.FAMILY></PERSON>

Contracted word groups

If two functions referring to but one person are linked by an *und/and* as in *der fr, here Warschauer KGB-Offizier und sowjetische Botschaftssekretär Wladimir Alganow* or *der CDU-Vorsitzende und Bundeskanzler, Helmut Kohl*, they are marked separately, but stay within the PERSON tag.

<PERSON><FUNCTION>Prime Minister</FUNCTION>&<FUNCTION>Chancellor of the Duchy of
Lancaster</FUNCTION> <NAME.GIVEN> Tony</NAME.GIVEN>
<NAME.FAMILY>Blair</NAME.FAMILY></PERSON >
<PERSON><FUNCTION>fr, here Warschauer KGB-Offizier</FUNCTION>&<FUNCTION>sowjetische
Botschaftssekretär</FUNCTION><NAME.GIVEN> Wladimir</NAME.GIVEN>
<NAME.FAMILY>Alganow</NAME.FAMILY></PERSON >

The reverse case, where one function refers to several persons, as in *die Minister Kinkel und Seehofer*, the different persons must be identified as such. For the second person, however, the semantic relation to the preceding FUNCTION is lost.

<PERSON><FUNCTION>Secretaries of State</FUNCTION> <NAME.FAMILY>Hurd</NAME.FAMILY>
and <NAME.FAMILY>Howard</NAME.FAMILY></PERSON>
<PERSON><FUNCTION>Chancellors</FUNCTION> <NAME.FAMILY>Lamont</NAME.FAMILY> and
<NAME.FAMILY>Clarke</NAME.FAMILY></PERSON>
<PERSON><FUNCTION>Minister</FUNCTION> <NAME.FAMILY>Kinkel</NAME.FAMILY></PERSON>
und <PERSON><NAME.FAMILY>Seehofer</NAME.FAMILY></PERSON>

ORGANISATION <ORG><ORGID></ORGID></ORG>

The ORG tag serves to identify organisational names and may occur anywhere in the running text. An inner tag (ORGID) allows to add special widely agreed-upon codes, such as codes from the Standard Industry Classification (SIC) list.

Tag content

The ORG tag typically contains the proper name of an organised group of people from all areas, including political, administrative, commercial, military, clerical, sports organisations (*Bank of England, Singapore Telecommunications Ltd; Bundesanstalt für Angestellte, FC Bayern München*).

Environmentalists yesterday hailed a decision by the <ORG>World Bank</ORG>

The <ORG>Centre for International Environmental Law</ORG> credited Mr. James Wolfensohn

A company's legal form (*e.g. Bro., Bros, Co., Corp., Inc., Ltd.; AG, KG, GmbH*) is considered part of its name and included in the ORG tag:

<ORG>Singapore Telecommunications Ltd</ORG><ORG> Reuters Limited</ORG>, <ORG> International Business Machines Corp</ORG>

<ORG>Langenscheidt KG</ORG>, <ORG>Jadekost GmbH & Co KG</ORG>, <ORG>KPM-Königliche Porzellan-Manufaktur GmbH</ORG>, <ORG>Schwartauer Werke GmbH & Co</ORG>

An abbreviation, following the organisational name directly, is included in the same ORG tag:

<ORG>Australian Securities Commission (ASC)</ORG>

<ORG>Singapore Telecommunications Ltd (SingTel)</ORG>

<ORG>Bundesanstalt für Angestellte (BfA)</ORG>, <ORG>Metallgesellschaft (MG)</ORG>

If the legal form is extended by an abbreviation, both get included in the ORG tag.

<ORG>Singapore Telecommunications Ltd (SingTel)</ORG>

<ORG>Metallgesellschaft AG (MG)</ORG>

In the course of a news message, a number of different synonymous expressions, such as acronyms, abbreviations, short forms (often the broader term), or paraphrases indicating the type of the organisation or the transliteration of a foreign name often refer to the same organisation. All these expressions are tagged:

<ORG>Bank of England</ORG>, the <ORG> Bank</ORG>

<ORG>Ministry of Agriculture, Fisheries and Food </ORG>, <ORG>Ministry of Agriculture</ORG>

<ORG>Australian Securities Commission (ASC)</ORG>, the <ORG>ASC</ORG>...

<ORG>Bundesanstalt für Angestellte (BfA)</ORG>, <ORG>Bundesanstalt für Angestellte</ORG>, <ORG>Bundesanstalt </ORG>, die <ORG>BfA</ORG>

der <ORG>russische Geheimdienst</ORG>, der <ORG>KGB</ORG>

It also includes personification of organisations, as in phrases, such as: *the Government said*, although this is nothing but a short form for the *spokesman or representative of the Government said*.

As an inner element the ORG tag may have an ORGID tag, which allows filling in special widely agreed upon identification codes which appear as attribute values of the nested element ORGID. In the following example, the codes from the Standard Industry Classification (SIC) list are applied.

<ORG ><ORGID IDSRC = SIC, VALUE=372> Airbus Industrie </ORG>

Specificity

City or country premodifying organisational names: In English, a noun modifier in front of an organisation may be the proper name of a city or country, referring to the location where the organisation is situated or the nationality it belongs to. A preceding adjective composed of a city or country and a hyphenated -based (as in *London-based*) has the same function, so does the combination with a Genitive (*London's*). German, instead, uses an adjective derived from the city's name (*Londoner*). It is recommended to include these pre modifiers in the ORG tag.

the <ORG> Winchester Crown Court</ORG>
 <ORG> Norfolk-based Lotus</ORG>
 <ORG> UK's Group Lotus</ORG>,
 <ORG> Britain's Labour party</ORG>
 die <ORG> Nürnbergberger Bundesanstalt für Arbeit</ORG>

Quite frequently the organisation type precedes the organisation name. If not tagged separately the valuable information given with this qualifier is lost.

the diversified utility group <ORG>VIAG AG</ORG>
 die Leichtmetall-Gießerei <ORG>Rackwitz GmbH</ORG>
 ihre Tochtergesellschaft <ORG>Malile S.A.</ORG>

Suggested modification to the element: Introduce new optional nested element OrgType to allow more precision.

<ORG><ORG.TYPE>drug manufacturer</ORG.TYPE>Wellcome</ORG>
 <ORG><ORG.TYPE>US-Rüstungskonzern</ORG.TYPE>McDonnell Douglas Corporation</ORG>

Locations postmodifying organisational names: An organisation is sometimes followed by a location. Depending on its function, the location must either be included in the ORG tag or it must be left out. Analogous to the preceding adjectives, the modifier that is part of the name or contributes to the precise identification of the organisation must be included.

<ORG> Anglo American Corp of South Africa</ORG>
 <ORG>Nomura International in London</ORG>
 <ORG>Arbeitsamt in Mannheim</ORG>

If the location is added for the reader's additional information only, it should be separately marked as LOCATION and thus as a serial relation. This is true for the following two examples:

<ORG>Brown University</ORG> in <LOCATION<CITY>Providence</CITY>, <STATE>R.I.
 </STATE></LOCATION>
 <ORG>TrizecHahn</ORG>, based in <LOCATION<CITY>Toronto</CITY></LOCATION>
 <ORG>Bundesanstalt für Meeresforschung</ORG> in <LOCATION<CITY>Hamburg</CITY>
 </LOCATION>

At times, it is hard to decide whether the location should be included in the ORG tag or not, as in *Nato-Militärausschuß in Brüssel*. If in doubt, the LOCATION tag should be used.

<ORG>Nato-Militärausschuß</ORG> in <LOCATION<CITY>Brüssel</CITY></LOCATION>

Organisation types, apart from preceding organisation names, may also follow them. They are tagged separately.

<ORG>Standard Life</ORG>, one of Europe's largest <ORG>insurers</ORG>

Notes: Alternatively, for greater precision and if ORGTYPE is permitted:

<ORG>Standard Life</ORG>, one of Europe's largest <ORGTYP>insurers</ORGTYP>
<ORG>Eagle Star</ORG>, one of <ORGTYP>BAT's insurance companies</ORGTYP>
<ORG>T&N, the <ORGTYP>UK automotive components company</ORGTYP></ORG>

If organisations are an integral part of functions, they are not marked separately, but are included within the FUNCTION tag. They may either precede or follow the headword.

<FUNCTION>Pakistan Muslim League (PML) leader</FUNCTION>
<FUNCTION>director of Schroder Unit Trusts</FUNCTION>
<FUNCTION>SPD-Fraktionsvorsitzender</FUNCTION>
<FUNCTION>Vorsitzender des rheinland-pfälzischen Gaststättenvereins</FUNCTION>

Multiple categorisation

At times, an expression categorised as an organisation name in one context may occur as a member of a different category in another, i.e. it allows for multiple semantic categorisation. This kind of ambiguity is highly context-dependent, and the decision for one or the other category is sometimes hard to make.

Typical cases of multiple categorisation with organisations are:

organisation | product: A product carries the company name by which it was produced, it should then be coded as an OBJECT.TITLE:

<ORG>Rover</ORG> was recently bought by <ORG>BMW</ORG>.
The new <OBJECT.TITLE>Rover</OBJECT.TITLE> 400 Series vehicles
<ORG>VW</ORG> will keine Mitarbeiter entlassen.
Ein <OBJECT.TITLE>VW</OBJECT.TITLE> läuft und läuft und läuft.

organisation | person(s): Organisations that carry the family name or the full name of their founders should be tagged as organisations:

<ORG> Schroder Unit Trusts</ORG>, [powerful] fund managers and unit trust providers such as
<ORG>Schroders</ORG>, <ORG>Adolf Darbo</ORG>
<ORG>Langenscheidt</ORG>hat weltweit 1 300 Mitarbeiter.
<ORG>Hugo Boss</ORG>hat wieder zugelegt.

Political parties may alternately be referred to by their organisational name or by addressing their members collectively: *The Conservative Party vs. the Tories, the Loughborough Conservatives; Freie Demokratische Partei, FDP vs. die Liberalen, die Freien Demokraten*. However, in some context, if members are addressed they merely refer to some specific individuals of the organisation and are not at all used synonymously to the whole. To be on the safe side, we suggest surrounding the organisation as well as the members, whether addressed collectively, as a subset or as a group of individuals, with an ORG tag.

the<ORG>Tories</ORG>, the <ORG>Conservatives</ORG>
The joint statement by <ORG>Labour</ORG> and the <ORG>Liberal Democrats</ORG>
<ORG>Freie Demokratische Partei (FDP)</ORG>, <ORG>Freie Demokratische Partei</ORG>,
<ORG>FDP</ORG>
die<ORG>Liberalen</ORG>, die <ORG>Freien Demokraten</ORG>

Note: If only one person (the Tory, der Grüne) or a small number of persons are addressed the FUNCTION tag should be used.

Some of the <FUNCTION> Liberal Democrats</FUNCTION>, however,

Dagegen sind einige <FUNCTION>Freien Demokraten</FUNCTION> der Meinung

organisation | location: To use a location figuratively when actually addressing an organisation is a productive language feature in English and German. Thus, the British Cabinet Office is referred to by *Downing Street*, the American President's Office by the *White House*, and the German Government by *Bonn*.

Locations that are used figuratively do not need any introduction, if the figurative meaning is well established:

<ORG>Wall Street</ORG>, <ORG>White House</ORG >

<ORG> University</ORG> victory after two-year battle with <ORG>Downing Street</ORG>

<ORG> Hardthöhe </ORG>, <ORG> Pentagon </ORG>,<ORG>Weißes Haus </ORG>

Note: A non-figurative use of *Downing Street* is applied in the following figure caption:

Douglas Hogg, the Agriculture Minister, who broke his ankle at home at the weekend, leaving
<LOCATION>Downing Street</LOCATION> yesterday

A country may function as an organisation in certain situations: *Deutschland* as a member state of the NATO clearly is an organisation, as a place where the river Rhine flows it is a COUNTRY. For consistency's sake, we suggest always coding a country as COUNTRY (cf. LOCATION).

Using the name of cities as a convenient short form instead of an official organisation name is a typical stylistic feature of the sports section, e.g. in a sentence like *Bremen gewann gegen Hamburger SV*, *Bremen* is a substitute for the football (soccer) club *SV Werder Bremen*. This may, however, also occur in other contexts, such as business news.

If a location is used as a short form for an organisation, the first mention of the organisation within the message will be its more explicit form. Organisations thus introduced may be coded as ORG.

<ORG>Halifax Building Society</ORG> ... <ORG>Halifax</ORG>

<ORG>Singapore Telecommunications Ltd (SingTel)</ORG> ... <ORG>Singapore</ORG>

<ORG>FC Bayern München </ORG> ... <ORG> Bayern </ORG>

If no organisational name is given in the news story as a key, the LOCATION tag should be used.

<LOCATION><CITY>Bremen</CITY></LOCATION> gewann gegen <ORG>Hamburger SV</ORG>

The question whether to tag a text unit as ORG or LOCATION may arise, e.g., with prisons, hospitals, hotels, air ports, theatres, museums. When *Honecker* left for Chile departing from the *Rhein-Main-Flughafen*, the airport is clearly addressed as a LOCATION. When somebody leaves the *Untersuchungshaftanstalt Frankfurt-Preungesheim*, it may be hard to decide whether to code one or the other.

As a rule, we code static structures like airports, prisons, hospitals, etc. as LOCATIONS:

The recent fire at <LOCATION>Heathrow</LOCATION> airport

Drei Wochen später konnte er die Untersuchungshaftanstalt <LOCATION>Frankfurt-Preungesheim</LOCATION> wieder verlassen.

Contracted word groups

Company types are often contracted into one to express the multi functionality of the type, as, for instance, in:

one of the largest <ORG> office building and shopping mall companies</ORG>

der <ORG >Rüstungs- und Luftfahrtkonzern McDonnell Douglas Corporation</ORG>

In the following example, however, two different organisation types are addressed and coded as such:

powerful <ORG >fund managers</ORG> and <ORG >unit trust providers</ORG>

Note: As already suggested, it would be more precise to code above examples as ORG.TYPE.

LOCATION <LOCATON> ... </LOCATION>

Tag content

The LOCATION tag identifies geographic locations and significant places (NITF, 8.46), it either contains mere text or structured information. The possible inclusions are: SUBLOCATION, CITY, REGION, STATE, COUNTRY, which subcategorise the geographic locations into a sequence of geo-political or administrative units. The subcategories may be used in combination with each other; in particular, to capture sequences of place names that are arranged in ascending order according to their size or administrative importance, such as *Cambridge, MA*. Whereas CITY, REGION, STATE, COUNTRY may well be used by themselves, although always within a LOCATION tag, SUBLOCATION is reserved to denote a relative position of a smaller LOCATION, i.e. a named part of a CITY or STATE. This would be marked up together and when the encompassing location occurs within the same tag, e.g.

<LOCATION><SUBLOCATION >Red Square</SUBLOCATION in<CITY> Moscow</CITY></LOCATION>

Note: In principle, locations have different functions within a news article: they can occur as independent semantic units of direct interest, but often are used as modifiers to other entities like persons, functions, organisations and events.

LOCATION <LOCATION></LOCATION>

Tag content

Places that are outside this range of the defined LOCATION subcategories, but may also be of interest, include a broad spectrum of geographical features, such as bay, island, isthmus, lake, landmass, mountain, ocean, peninsula, plateau, river, strait, waterfall, volcano, continent. They may also comprise significant man-made structures, such as famous buildings and constructions, bridges, walls, buildings, highways. For these the LOCATION tag without any inner elements may be used.

Passengers were offered crossings through the <LOCATION>Channel Tunnel</LOCATION>

a second fire had broken out in the <LOCATION>Iranian encampment on Mt Arafat</LOCATION>

Die höchste Erhebung des Harz, der <LOCATION>Brocken</LOCATION>

Suggested modification to the element: Introduce a LOCATION.TYPE to get structures, such as

die <LOCATION.TYPE>höchste Erhebung des Harz</LOCATION.TYPE>, der
<LOCATION>Brocken</LOCATION>

Multiple categorisation

Notes: As already mentioned, locations tend to be used figuratively. Countries and cities are often used when actually the government of a specific country or the government situated in a given city is meant (cf. ORG & location).

<ORG> University</ORG> victory after two-year battle with <ORG>Downing Street</ORG>

<ORG>Bonn</ORG> setzt sich durch.

At times, a location's name is conveniently used to address a specific event that took place there. Such instances should be tagged as `EVENTs` rather than as `LOCATIONs` (cf. event | location).

The consequences of `<EVENT>Chernobyl</EVENT>` are still felt today.

`<EVENT>Auschwitz</EVENT>` darf sich nicht wiederholen.

Static structures, such as architectural and engineering constructions, may be seen as `LOCATIONS` or addressed as `OBJECT.TITLES` because they can be considered works of art. A distinction is hard to make.

One thousand preachers are meeting in `<LOCATION>Westminster</LOCATION>` for congress.

Serial relations

More often than not, locations are further specified by other locations, either to help the reader's orientation or to clarify places whose name exists more than once. These serial relations are explicitly modelled in NITF. The way to code a relationship among locations is by surrounding a more complex expression by the `LOCATION` tag and to use its inner tags for differentiation purposes (e.g. `SUBLOCATION` with `CITY`, `REGION` or `STATE`, `CITY` with `STATE` or `COUNTRY`, `STATE` with `COUNTRY`).

Indicators often put in front of not-so-well-known locations are used for better identification, as in the following sentence: *town (of)* or *Cheshire town (of)*.

the highest concentration of millionaires is to be found in the `<LOCATION><STATE>Cheshire</STATE> town of <CITY>Alderley Edge</CITY></LOCATION>`

A slightly smaller number live north of `<LOCATION><CITY>Leeds </CITY> in the <REGION>Wetherby/Linton area</REGION></LOCATION>`

Der britische Brigadegeneral sagte auf der IFOR-Presskonferenz, die am Dienstag in der Serbenhochburg `<LOCATION><CITY>Pale</CITY> bei <CITY>Sarajevo</CITY></LOCATION>` abgehalten wurde

Note: Locations are frequently used as modifiers of organisations, functions, and events. See appropriate entries under `ORG`, `FUNCTION`, and `EVENT`.

Suggested modification to the element: Introduce `LOCATION.TYPE` as an inner element to the `LOCATION` element.

`SUBLOCATION <LOCATION><SUBLOCATION></SUBLOCATION></LOCATION>`

Tag content

A `SUBLOCATION` can only be used within the `LOCATION` tag and is meant to identify a small named geographic region or other entity, such as a neighbourhood or a park [cf. NITF, 8.68].

`<LOCATION><SUBLOCATION>Prince of Wales's garden</SUBLOCATION> in <STATE>Gloucestershire</STATE></LOCATION>`

outside the `<LOCATION><SUBLOCATION>St.Basil's Cathedral, Red Square,</SUBLOCATION <CITY>Moscow </CITY></LOCATION>`

Die Polizei ist sich zwar nicht sicher, ob er sich in einer Wohnung im Stadtteil `<LOCATION><SUBLOCATION>Hohenfelde</SUBLOCATION></LOCATION>` aufgehalten hat

In German, the names of cities may get an ending to indicate that a specific sublocation is situated there, districts are often connected to the city they belong to by a hyphen.

Wegen der doppelten Polizistenmorde auf dem `<LOCATION><CITY>Berliner </CITY><SUBLOCATION>B,lowplatz</SUBLOCATION></LOCATION>`

Während der Krawalle vor dem Asylbewerberheim in <LOCATION><SUBLOCATION>Rostock-Lichtenhagen</SUBLOCATION></LOCATION>

Multiple categorisation

Due to the fact that Ost-Berlin and West-Berlin were independent cities before 1989, they allow for multiple categorisation. As these denotations are quite often used now to indicate the Eastern or Western area of the city oblivious of an historic context, we suggest to tag them as SUBLOCATION.

<LOCATION><SUBLOCATION>Ost-Berlin</SUBLOCATION></LOCATION>

Serial relations

SUBLOCATION may be used in combination with the CITY, STATE, and possibly with REGION and COUNTRY tags.

<LOCATION><SUBLOCATION>Prince of Wales's garden</SUBLOCATION> in
<STATE>Gloucestershire</STATE></LOCATION>

CITY <LOCATION><CITY></CITY></LOCATION>

Tag content

As NITF does not offer different tags for settlements of different sizes, like cities, towns and villages, large and small administratively independent communities have to be marked up indiscriminately as CITY.

Born and raised in a working-class area of <LOCATION><CITY>Lima </CITY></LOCATION>

Western diplomats in <LOCATION><CITY>Moscow</CITY></LOCATION> are eager to scrutinise the wording of a "joint declaration"

Mr Blair told a rally of 1,400 people in <LOCATION><CITY>Stevenage</CITY></LOCATION>

Cities may be qualified in text as to their type; the generic description is not tagged:

the German resort of <LOCATION><CITY>Baden-Baden</CITY></LOCATION>

Rivers may identify Cities; often these rivers are part of the name.

<LOCATION><CITY>Stratford-on-Avon</CITY></LOCATION>

<LOCATION><CITY>Frankfurt am Main</CITY></LOCATION>

Paraphrases, if unambiguously referring to a unique city, are tagged as CITY.

<LOCATION><CITY>Spreemetropole</CITY></LOCATION>

<LOCATION><CITY>tschetschenische Hauptstadt</CITY></LOCATION>

Multiple categorisation

city | organisation: In a political context, multiple categorisation of cities occurs when the capital of a country is addressed figuratively instead of the administration or government of the country. Although an organisation is actually meant, the cities are tagged as such.

<LOCATION><CITY>Tehran</CITY></LOCATION> has a thriving nuclear co-operation programme with
<LOCATION><CITY>Moscow</CITY></LOCATION>

<ORG>Bundestag</ORG> und <ORG>Bundesregierung</ORG> müssen sich fügen, wenn
<LOCATION><CITY>Brüssel</CITY></LOCATION> die Tabakwerbung verbietet

In sport, it is common practice to name the place where a club is situated instead of the club's proper name, if the context is unambiguous. Again, we suggest tagging the cities as CITY, although, clearly, a sports club is addressed.

<LOCATION><CITY>Portsmouth</CITY></LOCATION>climbed four places to seventh

<LOCATION><CITY>Ipswich Town</CITY></LOCATION>virtually assured their place in the play-offs

<LOCATION><CITY>Bremen</CITY></LOCATION> schlug den <ORG>Hamburger SV</ORG> mit 4:1

city | state: In German, so-called *Stadtstaaten*, like *Berlin*, *Hamburg*, and *Bremen*, may be either referred to as a city or, less frequently, as a *Bundesland*. Because they are mostly mentioned as cities, they are coded as such.

<LOCATION><CITY>Bremen</CITY></LOCATION> profitiert von dem Länderausgleich

Note: Due to the fact that *Ost-Berlin (Ostberlin)* and *West-Berlin (Westberlin)* were independent cities before 1989, they allow for multiple categorisation. If these denotations are clearly used in a historic context, they may be tagged as CITY (cf. SUBLOCATION).

Die damalige Hauptstadt der DDR war <LOCATION><CITY>Ost-Berlin</CITY></LOCATION>

im <LOCATION><SUBLOCATION >Ostteil Berlins</SUBLOCATION></LOCATION> werden viele Gebäude restauriert

Serial relations

CITY may be used in combination with: SUBLOCATION, REGION, STATE, and COUNTRY.

Cities are also used to better identify other locations, e.g. buildings:

a three-day international congress on preaching at <LOCATION ><SUBLOCATION >Westminster Abbey</SUBLOCATION >, <CITY>London</CITY></LOCATION >

Note: Cities are used as modifiers or modifying adjectives to other concepts, e.g. function, person, event, and organisation. They are not separately tagged in these cases, but are included in the respective tags.

der die Analyse des <PERSON><FUNCTION>K`iner Doping-Fahnder</FUNCTION> Prof. <NAME.GIVEN>Manfred</NAME.GIVEN> <NAME.FAMILY>Donike</NAME.FAMILY></PERSON> ,berprüfen soll

Nach Aussage von <PERSON><FUNCTION>Rechtsanwalt</FUNCTION> <NAME.GIVEN>Peter</NAME.GIVEN><NAME.FAMILY>W`ssner</NAME.FAMILY>(Frankfurt)</PERSON >

REGION <LOCATION><REGION></REGION></LOCATION>

Tag content

We suggest tagging as REGION a named or paraphrased area of the earth (other than CITY, STATE, and COUNTRY) that is either geographically connected (*Balkans*, *Balkan*) or seen as a whole by some other characteristics (*Third World*, *Dritte Welt*), in particular, cultural, military or commercial criteria [cf. NITF, 8.6.1]. Continents and subcontinents should also be tagged as REGIONS.

they boarded a coach to the <LOCATION><REGION>Lake District</REGION></LOCATION>

<LOCATION><REGION>British waters</REGION></LOCATION> held 70 per cent of what became European fish

<LOCATION><REGION>Nato-Gebiet</REGION></LOCATION>

<LOCATION><REGION>Rhein und Ruhr</REGION></LOCATION>

<LOCATION><REGION>S,ddeutschland</REGION></LOCATION>

<LOCATION><REGION>Ex-Jugoslawien</REGION></LOCATION>

Note: Whereas mountain peaks (*Matterhorn*) and islands (*Guadeloupe, Corfu*) are tagged as LOCATION, mountain ranges (*Alps*) and island groups (*West Indies, fgäis*) are categorised as REGIONS.

Multiple categorisation

region | country: *Westdeutschland* and *Ostdeutschland*, which, at times, were informally used to refer to the states BRD and DDR before 1989, are now denoting regions within Germany. As in current usage, both indicate regions, we suggest tagging them as such.

Als die Wirtschaftwunderjahre <LOCATION><CITY>Westdeutschlands</CITY></LOCATION> vorbei waren

Serial relations

REGION may be used in combination with: SUBLOCATION, CITY, STATE, and COUNTRY.

Ein defekter Holzofen hat vermutlich das Feuer in einem Haus in
<LOCATION><CITY>Schopfheim</CITY> (<REGION>S,dbaden</REGION>)</LOCATION>ausgel'st

STATE <LOCATION><STATE></STATE></LOCATION>

Tag content

Compared to regions, which are geographical entities, states are geo-political entities on a lower level than countries, including provinces, counties in Britain, departements in France, and, in particular, the federal states of a country, e.g., the *Bundesländer* in Germany, the states in the U.S. [NITF, 8.67]. All the alternative names, i.e. names in any language (*Kalifornien, California*), short forms (*Bayern*), abbreviations (NRW) and post codes (*MA*), official (*Freistaat Bayern*) and non-official forms (*der Freistaat*) are tagged.

Opaque political situations may, of course, lead to inconsistencies in tagging.

<LOCATION><STATE>Virginia</STATE></LOCATION>

<LOCATION><STATE>Nordrhein-Westfalen</STATE></LOCATION>

Generic denotations, such as, *Departement, Bundesland, Grafschaft, US-Bundesstaat*, remain untagged (cf. Suggested modification: Introduction of LOCATION.TYPE).

wurde im US-Staat<LOCATION><STATE>Florida</STATE></LOCATION> verhaftet

Multiple categorisation

Note: In Germany, so called *Stadtstaaten*, like *Berlin, Hamburg*, and *Bremen*, are referred to either as a city or, less frequently, as a *Bundesland*. Because they are most frequently addressed as cities, we suggest tagging them as CITY, rather than as STATE.

Nächste Woche beginnt auch in <LOCATION><CITY>Berlin</CITY></LOCATION>,
<LOCATION><CITY>Bremen</CITY></LOCATION>,<LOCATION><STATE>Brandenburg</STATE></LO
CATION> und <LOCATION><STATE>Niedersachsen</STATE></LOCATION> der Unterricht wieder.

Serial relations

STATE may be used in combination with: SUBLOCATION, CITY, REGION, and COUNTRY.

Al Potter, 32, the helicopter captain, said when he returned to <LOCATION><SUBLOCATION>RNAS
Culdrose </SUBLOCATION>, <STATE>Cornwall</STATE></LOCATION>, that he had never seen such
bad conditions.

Nur beim letzten Wettbewerb in <LOCATION><CITY>Londover</CITY>
(<STATE>Maryland</STATE>)</LOCATION>landete die Deutsche auf Rang vier.

COUNTRY <LOCATION><COUNTRY></COUNTRY></LOCATION>

Tag content

The COUNTRY tag identifies internationally recognised/acknowledged geo-political units of a nation [cf. NITF, 8.18], such as *Great Britain*, the *United States of America*, *Japan*. All the alternative names, i.e. names in any language (*Schweiz*, *Switzerland*, *Suisse*), short forms (*Eidgenossenschaft*), abbreviations and international codes (*CH*), official (*Schweizerische Eidgenossenschaft*) and non-official forms are tagged.

In some of the worst weather since the Great Storm of 1987, hundreds of thousands of homes in
<LOCATION><COUNTRY>England</COUNTRY></LOCATION> were left without power.

Stürme legten die Fährverbindungen nach
<LOCATION><COUNTRY>England</COUNTRY></LOCATION> lahm

Multiple categorisation

country | organisation: Countries, in some contexts, function as organisations, if, for example, they are referred to as members of other organisations (NATO, UNO, etc.). We decided to code countries as such independent of the context.

Plans to bring <LOCATION><COUNTRY>Britain</COUNTRY></LOCATION> closer into line with other
<ORG>EuropeanUnion</ORG> countries

der Beitritt <LOCATION><COUNTRY>Rußlands</COUNTRY></LOCATION> zur <ORG>NATO</ORG>

Serial relations

COUNTRY may be used in combination with SUBLOCATION, CITY, REGION, and STATE.

VIRTUAL LOCATION <VIRTLOC></VIRTLOC>

Virtual locations, such as email or WWW addresses, are tagged using the VIRTLOC tag. VIRTLOC is no inner element of LOCATION.

The Times newspaper has launched a website at <VIRTLOC>http://www.the-times.co.uk</VIRTLOC>

EVENT <EVENT></EVENT>

Tag content

Marking up events [NITF, 8.28] should be limited to news-worthy events or events that carry news value in the sense of journalism. Factors of news value are, e.g. significance, proximity, prominence of the involved persons, consequence, unusualness, human interest, timeliness. The expected value for archival and long-term interest should also be considered.

Analogous to organisations, events may actually be differentiated into event types (*General Election*, *London Marathon*; *Flugzeugabsturz*) and into specific instances (*General Election 97*, *Atlanta Olympics*; *der TWA-Absturz*) that can be rooted in place and time. In the course of a message or

when the context is clear from the beginning, concrete events tend to be referred to by a synonymous abstracted denotation, often one of the broader terms: *the Olympics*, *Olympiade* instead of *the Atlanta Olympics 1996*, *die Olympischen Sommerspiele 1996 in Atlanta*.

Events get tagged, if there is an unambiguously identified event of news value, e.g., *Wimbledon 96*; *Golfkrieg*, *Fußball-Europameisterschaft 1996*. Their short synonymous expressions are also coded.

<EVENT>1997 Flora London Marathon</EVENT>, the <EVENT>Marathon</EVENT>
<EVENT>May 1 1997 general election</EVENT>, the <EVENT>election</EVENT>
<EVENT>Atlanta Olympics</EVENT>, the <EVENT>Olympics</EVENT>
<EVENT>Biennale 1997</EVENT>, die <EVENT>Biennale</EVENT>
<EVENT>Bundestagswahlen 1998</EVENT>, die <EVENT>Wahlen</EVENT>
<EVENT>Fußball-Europameisterschaft 1996</EVENT>, die <EVENT>EM</EVENT>

Further, events are tagged if they are re-occurring events and typical for the different news sections, such as, *general election*, *election rally*; *Parteitag*, *Landtagswahl* (politics), *European Football Championship*, *Olympics*; *EM*, *Olympiade* (sports), ... *Uraufführung*, *Premiere*, *Dokumenta*, *Biennale* (culture).

<EVENT>general election</EVENT>, <EVENT>referendum</EVENT>
<EVENT>Olympic Winter Games</EVENT>
<EVENT>hostage-taking</EVENT>
<EVENT>bombing</EVENT>
the Referendum Party staged an <EVENT>election rally</EVENT> at Alexandra Palace in London
the <EVENT>Haj</EVENT> has been dogged by tragedy
<EVENT>Parteitag</EVENT>
<EVENT>SPD-Parteitag</EVENT>

Specificity

As concrete events happen somewhere sometime, places and dates are used to identify them. Periodically re-occurring events may carry ordinal numbers. Places, dates and numbers are considered part of the name and therefore are not separately tagged. They may all be used as pre- or postmodifiers.

<EVENT>London Marathon</EVENT>, the <EVENT>1960 Olympics</EVENT>
der <EVENT> Mannheimer SPD-Parteitag</EVENT>
<EVENT>Landtagswahl in Baden-Württemberg</EVENT>
<EVENT>Dokumenta X</EVENT>, die <EVENT>Olympiade 1960</EVENT>

Multiple categorisation

event | location: At times, a location's name is conveniently used to address a specific event that took place there. Such instances should be tagged as EVENTS rather than as LOCATIONS.

The consequences of <EVENT>Chernobyl</EVENT> are still felt today.
<EVENT>Auschwitz</EVENT> darf sich nicht wiederholen.

Suggested modification: Introduce an EVENT.TYPE element.

TIME EXPRESSIONS <CHRON></CHRON>

Tag content

Concrete dates, such as the year of 1996, are tagged as CHRON.

<CHRON>1996</CHRON>

The days of the week, religious and bank holidays, and relative time expressions that may be attributed with a concrete date are also tagged. The date is inserted as an attribute value in the normalised form: YYYYMMDD.

On <CHRON NORM="19970502">Friday </CHRON> Tony Blair became Prime Minister

<CHRON NORM="19961216">Montag</CHRON>, <CHRON NORM="19961121">Buß- und Betttag</CHRON>, <CHRON NORM="19961215">gestern</CHRON>, <CHRON NORM="1995">letztes Jahr</CHRON>

MONETARY ITEMS <MONEY></MONEY>

Tag content

MONEY identifies monetary items. An attribute allows naming of the currency (type of unit) used.

a net fall in earnings per share of <MONEY UNIT=pence Sterling>10.2</MONEY>p this year, while the net asset value will fall by £<MONEY UNIT=pound Sterling>43 million</MONEY>

In the Var alone, where <NUM>20</NUM> communities were affected, the loss is estimated at <MONEY UNIT=French franc>80 million</MONEY> francs (£<MONEY UNIT=pound Sterling>8.8 million</MONEY>)

Holst hatte sich am vergangenen Montag mit <MONEY UNIT=DM>157 000</MONEY> Mark in der Tasche gestellt.

es entstand ein Sachschaden von mehreren <MONEY UNIT=DM >10 000</MONEY> Mark

Time intervals, such as die Wintersaison 1997/98 or 1931-1945 are also coded as CHRON.

NUMERIC DATA <NUM></NUM>

Tag content

Numeric data include all numbers, either stated as digits or characters, with the exception of MONEY items and CHRON elements denoting time and date. The NUM element may contain text or numbers, or may include structured information. Allowed elements are FRACTION, SUBSCRIPT, or SUPERSCRIPIT.

The worst [tragedy] was <CHRON NORM="1990">seven years ago</CHRON> when <NUM>1,426</NUM> pilgrims died

At <LOCATION>Gatwick</LOCATION>, which handles between <NUM>50</NUM> and <NUM>60</NUM> flights in and out every hour, more than <NUM>8,000</NUM> departing passengers had to be looked after

Rodney Barnes, <NUM>31</NUM>, who became known as the Mayfair rapist

Die Bewohner im Alter von <NUM>67</NUM> und <NUM>70</NUM> Jahren konnten das Anwesen rechtzeitig verlassen.

Bosnische Serben haben in den vergangenen <NUM>acht</NUM> Tagen angeblich <NUM>elf</NUM> moslemische Zivilisten verschleppt.

Special structures as 2:3 (results of a football game) are also tagged as NUM.

<NUM>2</NUM>: <NUM>3</NUM>

QUOTATION: QUOTE <Q></Q>, BLOCK QUOTE <BQ></BQ>

Tag content

According to the NITF comments, quote tags are used to either "encapsulate quoted information" (inline quote, 8.61) or to mark "material quoted from an external source" (block quote, 8.9). This means that indirect discourse remains unmarked and merely the exact rendering of a person's words or text cited from a written source get tagged as quotes. The two quote tag sets do not differentiate different quote contents, but different desired layouts. However, no information is given as to the expected length of a (shorter) inline quote compared to a (longer) block quote. Style guides as the Chicago Manual of Style (13th Edition) suggest to run shorter "quoted words, phrases and sentences" into the text enclosing them in quotation marks (p.289), whereas they advise to set off longer quoted matter, eight or ten lines, from the surrounding text by indenting them and thus creating a block quotation. HTML also makes a distinction between inline quotes and block quotations. Newspapers because of their column style hardly ever use the indented block quote, but rather delimit all quotes using quotation marks.

News stories of all sections may be interspersed with quotes. The markup of quotes implies problems different from the other more object-like semantic units, mainly because marking quotes by themselves is useless, if not at least originator or source of the quote are also made explicit. In fact, for long-term relevance a complete quote structure should be preserved consisting of the quoted text, the source, the recipient, the subject, the occasion, and the date. This, however, would mean a tremendous overhead, if the quoted text were irrelevant. To separate the goats from the sheep, one would have to consider the many reasons why journalists use quotes in their stories, which is not only for authenticity but also to add liveliness or to achieve distance to what was said. And the relevance of a quote would depend on the importance of the message its holds, its originator, the importance of the topic or occasion, etc.

NITF provides two different content models for quote and block quote. A short quote may contain character string or any of the inline elements. A block quote must contain a heading or a block and it may contain the source getting the credit for the quote.

INLINE QUOTE <Q></Q>

Inline quotes occur within running text. Quite often they are an integral part of a sentence, however offset by typographic means like quotation marks or italics. They may also be part of a quote structure consisting of the quote proper and a reporting clause that contains the source of the quotation and the reporting verb (e.g. *he said*). Reporting clauses may preface or follow the quotation, or they may interrupt a quote.

Integrated inline quote emphasising key phrases:

Mr Blair dubbed his programme <Q>both practical and radical</Q>

Quote preceded by reporting clause:

He added: <Q>We face the prospect of rising welfare bills, but combined with rising poverty and social division.</Q>

Quote followed by reporting clause:

<Q>... commands that Honourable Members... </Q> continued Black Rod.

Quote followed by reporting clause:

<Q>In Mr Mandelson's opinion,</Q> added the mordant Kaufman, <Q>total sycophancy must be regarded as a suspiciously lukewarm form of support.</Q>

Quotes in Quotes:

Ms Maher said of the letter: <Q>It doesn't mince words. It says <Q>Dear Madam</Q>, not even using the person's name. It also says the first payment is due within seven days. Then it gets nasty. </Q>

BLOCK QUOTE <QB></QB>

Block quotations may also occur with or without reporting clause. The block quote may stretch over one or several paragraphs. If the block quotes are formally introduced, the reporting clause and the first part of the block quote often form one paragraph. Interviews are sequences of BLOCK QUOTES, message types, such as "Im Wortlaut" (dpa) or legal documents consist may be considered one lengthy BLOCK QUOTE.

BLOCK QUOTES may start or end within a paragraph and thus will cause markup problems because SGML does not support overlapping structures.

Unprefaced multiple-paragraph block quotes:

<P>Mr Mugabe, however, yesterday compared his approach with that of the government of post-apartheid South Africa.</P>

<BQ><BLOCK><P>We want to ensure that in the process ...</P>

<P>The indigenuous people must have a stake ...</P>

<P>We want to turn workers into entrepreneurs ...</P></BLOCK><BQ>

Prefaced multiple-paragraph block quotes:

<P>Asked why Ferguson should choose to involve himself in an issue that did not affect his club, Leaver said:<Q>He likes it, that's what he does. Anyway, what did he expect us to do? Were we really going to ask the players in the other matches to sit around in the dressing room at half-time while Tottenham and Coventry played catch-up? Were all the spectators expected to mill around and the stewards expected to look after them for an indeterminate period? </Q></P>

<BQ><BLOCK><P>It seems to me that if you ask sensible questions, there could be only one sensible answer, and that was the decision that was taken.</P>

<P>They [Manchester United] feel that they should be given consideration that other clubs may not get, but I have to be even-handed and treat all clubs in the same way. If I fall out with Manchester United, it is because I'm trying to be fair. </P></BLOCK></BQ>

Note: It would be desirable to capture the complete quote as BQ.

Interrupted multiple-paragraph block quotes:

<P><Q>The coming era of digital personal communications is an era of converging technologies...,</Q> says Mr. Eckehard Pfeiffer, president and chief executive of Compaq Computer. </P>

<BQ><BLOCK><P>More and more, the computer, broadcast, cable, telephone, satellite, and media entertainment industries will find themselves part of a much larger market-place. ...

</P></BLOCK><CREDIT> Eckehard Pfeiffer</CREDIT></BQ></BLOCK>

<P>This example clearly shows by its topic, i.e. the convergence of media technology, that there is but one long quote. </P>

Note: It would be desirable to capture the complete quote as BQ.

Mixed inline and block quotes:

<P>The growth of the Internet ... has also encouraged the vision of access to information, <Q>any time, any place, anywhere</Q></P>.

<BQ><BLOCK><P>The Internet has gone a long way toward illustrating the demand for that access ...
</P></BLOCK></BQ>

Suggested modification to the element: The BLOCK QUOTE element has a recursive structure that allows repetition of the required sequence BQ/BLOCK endlessly. There is no way to mark a BLOCK QUOTE that starts in the middle of a paragraph, a principal problem with SGML structures that typically do not allow for overlapping elements.

Multiple categorisation

Note: Quoted text is subject to multiple categorisation only in a formal way: Apart from indicating a direct citation, quotation marks are also used to delimit, for instance, object names, ironic denotations, literal translations, euphemisms, neologisms, unfamiliar or exotic expressions. Those instances should not be coded as quotes, but surrounded by the OBJECT.TITLE tag or left unmarked.

Serial relations

Items belonging to the above-mentioned quote structure may be considered serial relations, if they occur in the close context of the quote. The person credited with a quote, for instance, may often occur in the reporting clause, as in

<PERSON>Mr <NAME.FAMILY>Blair</NAME.FAMILY></PERSON> made plain that <Q>we have reached the limits of the public's willingness to fund an unreformed welfare system through ever higher taxes</Q>.

There is no formal way of associating an inline quote with its source. Within a BLOCK QUOTE the CREDIT tag fulfils this function.

<BQ><BLOCK><P>More and more, the computer, broadcast, cable, telephone, satellite, and media entertainment industries will find themselves part of a much larger market-place. ...
</P></BLOCK><CREDIT> Eckehard Pfeiffer</CREDIT></BQ></BLOCK>

OBJECT.TITLE <OBJECT.TITLE></OBJECT.TITLE>

Tag content

The NITF documentation states that the OBJECT.TITLE tag is meant to contain the title of an object within inline text, including items such as books, songs, art etc. [8.53]. In other words, OBJECT.TITLES should include named news-relevant world objects as, for instance, publication and media types (book, newspaper, CD, TV series), mass media channels (TV channels, radio stations), titles of awards and prizes, names of products and product lines, art objects, animals, ships, buildings. It remains to be investigated whether news-relevant objects must be associated with an EVENT.

<OBJECT.TITLE>Reuters World Service</OBJECT.TITLE> is its flagship Media product.

A team spokesman said the "<OBJECT.TITLE>Solo Spirit</OBJECT.TITLE>" still had more fuel than at this point in last year's attempt.

"Spiel's noch einmal, Sam" - Ingrid Bergmans berühmte Aufforderung aus

"<OBJECT.TITLE>Casablanca</OBJECT.TITLE>" nahmen die TV-Anstalten noch nie so wörtlich wie 1997

Weiss übernahm später den Part des Kapitäns in der ZDF-Unterhaltungsreihe "<OBJECT.TITLE>Das Traumschiff</OBJECT.TITLE>"

Suggested modification to the element: In news stories, object names are often preceded or followed by a descriptive identifier, such as, the *drama* "Dantons Tod". We suggest to either tag this seperately as an Object.Type or to introduce an attribute Object.Type.