
NLP Interchange Format Documentation

Release 2.1-RC

Sebastian Hellmann Martin Brümmer Markus Ackermann

October 17, 2016

1	Provenance and Confidence for NIF annotations	3
1.1	Provenance and Confidence using Companion Properties	3
1.2	Using Generic Provenance and Confidence Properties	7

The NLP Interchange Format (NIF) is an RDF/OWL-based format that aims to achieve interoperability between Natural Language Processing (NLP) tools, language resources and annotations. NIF consists of specifications, ontologies and software. These documents describe the upcoming *NIF 2.1* release, currently under revision by the broader community.

Although the road to complete interoperability is still long, NIF is already successful in providing best practices and a solid foundation for the most frequent use cases. This foundation is created by:

- Reusing existing standards such as RDF, OWL 2, the PROV Ontology and LAF (ISO 24612)
- Utilisation of NIF identifiers in the Internationalization Tag Set (ITS) Version 2.0
- Royalty-free publishing of all parts of NIF under an open license.

NIF comprises a set of RDF vocabularies and ontologies, which have stable identifiers, persistent hosting, an open license and a community approved meaning.

NIF is driven by its open and welcome-to-join community project NLP2RDF, consisting of a mailing list, a GitHub Project and a blog web site.

Contents:

Provenance and Confidence for NIF annotations

NIF 2.1 introduces additional vocabulary to express provenance and confidence information for annotations. This section will present two possible approaches to assign provenance and confidence information to annotations:

- a more compact representation using previously introduced or provided *companion properties*
- a simpler, but more verbose representation using the generic `nif:provenance` and `nif:confidence` properties

For a running example, assume the following short example sentence:

```
@prefix nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#> .

@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix ex: <http://example.org/nif-21/acq#> .

ex:doc_offset_0_52
  a nif:OffsetBasedString, nif:Context ;
  nif:beginIndex "0"^^xsd:nonNegativeInteger ;
  nif:endIndex "52"^^xsd:nonNegativeInteger ;
  nif:anchorOf "Apple acquired Metaio, an Augmented Reality company."^^xsd:string .
```

1.1 Provenance and Confidence using Companion Properties

1.1.1 Entity Spotting and Linking

Sending this NIF document to an entity spotting and linking service, e.g. possibly a future revision of the [FREMEE-Entity DBpedia Spotlight Service](#) could yield and RDF result similar to:

```
@prefix nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#> .
@prefix itsrdf: <http://www.w3.org/2005/11/its/rdf#> .
@prefix dbp: <http://dbpedia.org/resource/> .
@prefix dbpo: <http://dbpedia.org/ontology/> .
@prefix nerd: <http://nerd.eurecom.fr/ontology#> .
@prefix freme-api: <http://api.freme-project.eu/example/> .

@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix ex: <http://example.org/nif-21/acq#> .

[] a owl:Ontology ;
  a nif:OffsetBasedString, nif:Context ;
```

```

nif:beginIndex "0"^^xsd:nonNegativeInteger ;
nif:endIndex "52"^^xsd:nonNegativeInteger ;
nif:isString "Apple acquired Metaio, an Augmented Reality company."^^xsd:string .

ex:doc_offset_0_5
  a nif:OffsetBasedString ;
  nif:beginIndex "0"^^xsd:nonNegativeInteger ;
  nif:endIndex "5"^^xsd:nonNegativeInteger ;
  nif:anchorOf "Apple"^^xsd:string ;
  nif:referenceContext ex:doc_offset_0_52 ;
  # entity spotting information
  a nif:EntityOccurrence ;
  nif:entityOccurrenceConf "0.7"^^xsd:decimal ;
  nif:entityOccurrenceProv freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments ;
  # primary entity linking result
  itsrdf:taIdentRef dbp:Apple_Inc ;
  nif:taIdentConf "0.8"^^xsd:decimal ;
  nif:taIdentProv freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments ;
  itsrdf:taClassRef nerd:Organization ;
  nif:taClassConf "0.95"^^xsd:decimal ;
  nif:taClassProv freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments ;
  # alternative, less probable entity linking result
  nif:annotationUnit [
    itsrdf:taIdentRef dbp:Apple_Bank_for_Savings ;
    nif:taIdentConf "0.3"^^xsd:decimal ;
    nif:taIdentProv freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments ;
    itsrdf:taClassRef dbpo:Bank ;
    nif:taClassConf "0.9"^^xsd:decimal ;
    nif:taClassProv freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments ;
  ] .

ex:doc_offset_15_21
  a nif:OffsetBasedString ;
  nif:beginIndex "15"^^xsd:nonNegativeInteger ;
  nif:endIndex "21"^^xsd:nonNegativeInteger ;
  nif:anchorOf "Metaio"^^xsd:string ;
  nif:referenceContext ex:doc_offset_0_52 ;
  # entity spotting information
  a nif:EntityOccurrence ;
  nif:entityOccurrenceConf "0.95"^^xsd:decimal ;
  nif:entityOccurrenceProv freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments ;
  # entity linking result
  itsrdf:taIdentRef dbp:Metaio_GmbH ;
  nif:taIdentConf "0.9"^^xsd:decimal ;
  nif:taIdentProv freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments ;
  itsrdf:taClassRef dbpo:Company ;
  nif:taClassConf "0.85"^^xsd:decimal ;
  nif:taClassProv freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments .

```

Note: The current implementations of Freme services do not produce data as described here, thus also example is used instead of an acutally valid API version number. The discussed RDF data should rather be interpreted as suggestion/basis for discussion.

The service introduced two NIF substring resources that were spotted as potential named entities. Each substring resource carries several pieces of annotation information:

spotting information The mere fact that a certain substring has been identified as a (likely) reference to a named entity. This expressed in NIF 2.1 by assigning the `nif:EntityOccurrence` class to the substring resource.

entity linking information (Candidate) references to Linked Data identifiers for mentioned named entities or classification or referenced entities into one or several categories. For referencing, the `itsrdf:taIdentRef` property from [ITSRDF](#) is used.

According to the *companion properties* approach, for each of the used annotating properties and `nif:TextSpanAnnotation` subclasses a pair of specific and related subproperties of `nif:provenance` and `nif:confidence` were introduced:

annotation property/class	provenance property	confidence property
<code>nif:EntityOccurrence</code>	<code>nif:entityOccurrenceProv</code>	<code>nif:entityOccurrenceConf</code>
<code>itsrdf:taIdentRef</code>	<code>nif:taIdentProv</code>	<code>nif:taIdentConf</code>

Provenance properties reference either `prov:Agent` or `prov:Activity` resources providing details on either just the annotator (be it man or machine) or also additionally on the annotation process. An outline for an agent description for our example:

```
@prefix doap: <http://usefulinc.com/ns/doap#>
@prefix prov: <http://www.w3.org/ns/prov#>
@prefix freme: <http://freme-project.eu/example/>
@prefix freme-api: <http://api.freme-project.eu/example>
```

```
freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments
  a prov:SoftwareAgent, doap:Version ;
  doap:shortdesc "NIF REST API for entity recognition and linking using DBpedia Spotlight engine" ;
  doap:revision "0.x (example)" .
# [...]
```

```
freme:description#project
  a doap:Project ;
  doap:release freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments ;
  doap:vendor freme:description#consortium .
# [...]
```

Confidence properties provide a numeric measure for the degree of certainty of annotating agent when assigning the annotation as a rational number between 0 and 1.

Obviously only one property assertion for a specific companion property can be made for the same `nif:String` resource without causing ambiguity. Thus, whenever several alternative annotations on the same aspect are to be expressed, additional `nif:AnnotationUnit` resources can be created and linked to the `nif:String` annotated resource. In the example, such an `nif:AnnotationUnit` is used for the alternative, less probable entity link for string `ex:doc_offset_0_5` to `dbp:Apple_Bank_for_Savings`. The same `nif:AnnotationUnit` resource can be (re-)used to host multiple annotation statements with provenance and confidence via companion properties, as long as unequivocalness is ensured.

To ensure possibilities to validate such unambiguity and to ensure that provenance and confidence information using companion properties is completely machine actionable, explicit links between companion properties and their corresponding annotating vocabulary items, as in this excerpt of the current NIF 2.1 Core ontology draft:

```
nif:EntityOccurrence nif:confidenceProperty nif:entityOccurrenceConf ;
  nif:provenanceProperty nif:entityOccurrenceProv .

nif:TermOccurrence nif:confidenceProperty nif:termOccurrenceConf ;
  nif:provenanceProperty nif:termOccurrenceProv .

itsrdf:taIdentRef nif:confidenceProperty nif:taIdentConf ;
  nif:provenanceProperty nif:taIdentProv .
```

```
itsrdf:taClassRef nif:confidenceProperty nif:taClassConf ;
nif:provenanceProperty nif:taClassProv .
```

1.1.2 Terminology Annotation

In a similar way as presented for named entities information about term recognition and referencing can be provided, by APIs like the [FREME e-Terminology Service](#) a possible addition to the result listing in *the result listing of the previous section* could be:

```
a nif:OffsetBasedString ;
nif:beginIndex "26"^^xsd:nonNegativeInteger ;
nif:endIndex "43"^^xsd:nonNegativeInteger ;
nif:anchorOf "Augmented Reality"^^xsd:string ;
nif:referenceContext ex:doc_offset_0_52 ;
# entity spotting information
a nif:TermOccurrence ;
nif:termOccurrenceConf "0.7"^^xsd:decimal ;
nif:termOccurrenceProv fremen-api:description%2Fe-terminology%2Ftilde ;
# term linking result
itsrdf:termInfoRef <https://term.tilde.com/terms/998795> ;
nif:termInfoConf "0.65"^^xsd:decimal ;
nif:termInfoProv fremen-api:description%2Fe-terminology%2Ftilde .
```

1.1.3 Relation of NIF 2.1. companion properties to ITS RDF properties

`itsrdf:taConfidence` is very similar to both `nif-ann:taIdentConf` and `nif-ann:taClassConf`, but is specified to provide a common confidence value for both the link to a concrete entity reference and an entity type associated with this entity. Since NIF 2.1 also wanted to be able to express e.g. output of general entity spotters that also assign classes to spotted entities, but are unable to conclusively disambiguate them¹, NIF introduced its own specialised properties.

In cases when NIF 2.1 is actually to be used to describe term linking output from tools in line with the ITS premises, `itsrdf:taConfidence` and the corresponding `itsrdf:taAnnotatorRef` can be used alternatively to the NIF 2.1 companion properties. The provenance reference for `itsrdf:taAnnotatorRef` still should be either `prov:Agent` or `prov:Activity`

```
@prefix nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#> .
@prefix nif-ann: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-annotation#> .
@prefix itsrdf: <http://www.w3.org/2005/11/its/rdf#> .
@prefix dbp: <http://dbpedia.org/resource/> .
@prefix dbpo: <http://dbpedia.org/ontology/> .
@prefix nerd: <http://nerd.eurecom.fr/ontology#> .

@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix ex: <http://example.org/nif-21/dresden#> .

ex:doc_offset_0_7
  a nif:OffsetBasedString ;
  nif:beginIndex "0"^^xsd:nonNegativeInteger ;
  nif:endIndex "7"^^xsd:nonNegativeInteger ;
  nif:anchorOf "Dresden"^^xsd:string ;
  nif:referenceContext ex:doc_offset_0_62 ;
```

¹ Think for example for a simple gazeteer-based spotting service. It can easily spot ‘Dresden’ and might contain type data associating occurrences of this string with the category ‘populated place’, whilst lacking logic to hazard an informed guess whether it’s the city with that name in Germany, the United Kingdom, the US or Canada.

```

a nif-ann:EntityOccurrence ;
nif-ann:entityOccurrenceConf "0.88"^^xsd:decimal ;
nif-ann:entityOccurrenceProv ex:simple-spotter-service ;
nif-ann:annotationUnit [
  itsrdf:taIdentRef dbp:Dresden ;
  itsrdf:taClassRef dbo:City ;
  itsrdf:taConfidence "0.80"^^xsd:decimal ;
  itsrdf:taAnnotatorRef ex:linker-service
] .

```

1.2 Using Generic Provenance and Confidence Properties

Usage of compation properties allows to offer a default value for each annotation aspect and allows two reduce the number of `nif:AnnotationUnit` resources that must be synthesized to prevent ambiguities. However, they also increase technical complexity for consumption of provenance and confidence information. Using exclusively the generic `nif:provenance` and `nif:confidence` properties directly simplifies generation and consumption of this information, at the cost of additional RDF resources required to express equivalent data. Using only these generic properties to express the same annotation as discussed in *Provenance and Confidence using Companion Properties*:

```

@prefix nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#> .
@prefix itsrdf: <http://www.w3.org/2005/11/its/rdf#> .
@prefix dbp: <http://dbpedia.org/resource/> .
@prefix dbpo: <http://dbpedia.org/ontology/> .
@prefix nerd: <http://nerd.eurecom.fr/ontology#> .
@prefix freme-api: <http://api.freme-project.eu/example/> .

@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix ex: <http://example.org/nif-21/acq#> .

[[] a owl:Ontology ;
  a nif:OffsetBasedString, nif:Context ;
  nif:beginIndex "0"^^xsd:nonNegativeInteger ;
  nif:endIndex "52"^^xsd:nonNegativeInteger ;
  nif:isString "Apple acquired Metaio, an Augmented Reality company."^^xsd:string .

ex:doc_offset_0_5
  a nif:OffsetBasedString ;
  nif:beginIndex "0"^^xsd:nonNegativeInteger ;
  nif:endIndex "5"^^xsd:nonNegativeInteger ;
  nif:anchorOf "Apple"^^xsd:string ;
  nif:referenceContext ex:doc_offset_0_52 ;
  # entity spotting information
  nif:annotationUnit [
    a nif:EntityOccurrence ;
    nif:confidence "0.7"^^xsd:decimal ;
    nif:provenance freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments ;
  ] ;
  # primary entity linking result
  nif:annotationUnit [
    itsrdf:taIdentRef dbp:Apple_Inc ;
    nif:confidence "0.8"^^xsd:decimal ;
    nif:provenance freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments ;
  ] ;
  nif:annotationUnit [
    itsrdf:taClassRef nerd:Organization ;

```

```
nif:confidence "0.95"^^xsd:decimal ;
nif:provenance freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments ;
] ;
# alternative, less probable entity linking result
nif:annotationUnit [
  itsrdf:taIdentRef dbp:Apple_Bank_for_Savings ;
  nif:confidence "0.3"^^xsd:decimal ;
  nif:provenance freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments
] ;
nif:annotationUnit [
  itsrdf:taClassRef dbpo:Bank ;
  nif:confidence "0.9"^^xsd:decimal ;
  nif:provenance freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments
] .

ex:doc_offset_15_21
a nif:OffsetBasedString ;
nif:beginIndex "15"^^xsd:nonNegativeInteger ;
nif:endIndex "21"^^xsd:nonNegativeInteger ;
nif:anchorOf "Metaio"^^xsd:string ;
nif:referenceContext ex:doc_offset_0_52 ;
# entity spotting information
a nif:EntityOccurrence ;
nif:confidence "0.95"^^xsd:decimal ;
nif:provenance freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments ;
# entity linking result
nif:annotationUnit [
  itsrdf:taIdentRef dbp:Metaio_GmbH ;
  nif:confidence "0.9"^^xsd:decimal ;
  nif:provenance freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments ;
] ;
nif:annotationUnit [
  itsrdf:taClassRef dbpo:Company ;
  nif:confidence "0.85"^^xsd:decimal ;
  nif:provenance freme-api:description%2Fe-entity%2Fdbpedia-spotlight%2Fdocuments ;
] .

ex:doc_offset_26_43
a nif:OffsetBasedString ;
nif:beginIndex "26"^^xsd:nonNegativeInteger ;
nif:endIndex "43"^^xsd:nonNegativeInteger ;
nif:anchorOf "Augmented Reality"^^xsd:string ;
nif:referenceContext ex:doc_offset_0_52 ;
# entity spotting information
a nif:TermOccurrence ;
nif:confidence "0.7"^^xsd:decimal ;
nif:provenance freme-api:description%2Fe-terminology%2Ftilde ;
# term linking result
nif:annotationUnit [
  itsrdf:termInfoRef <https://term.tilde.com/terms/998795> ;
  nif:confidence "0.65"^^xsd:decimal ;
  nif:provenance freme-api:description%2Fe-terminology%2Ftilde ;
] .
```

Note: `nif:confidence` and `nif:provenance` can only be attached to `nif:AnnotationUnit` instances, not to `nif:String` instances directly.

- `genindex`