



Rupert Westenthaler

APACHECON
EUROPE

Open Annotation Support for
Apache Stanbol

Apache Stanbol Enhancer

```
curl -X POST -H "Accept: text/turtle" -H "Content-type: text/plain" \  
  --data "The Stanbol enhancer can detect famous cities such as \  
  Paris and people such as Bob Marley." \  
  http://localhost:8080/enhancer
```

**POST
content**

Enhancement Chain: **default** all 5 engines available

- 🔧 **tika** (optional , TikaEngine)
- 🔧 **langid** (required , LangIdEnhancementEngine)
- 🔧 **ner** (required , NamedEntityExtractionEnhancementEngine)
- 🔧 **dbpediaLinking** (required , NamedEntityTaggingEngine)

**Analysis
Chain**

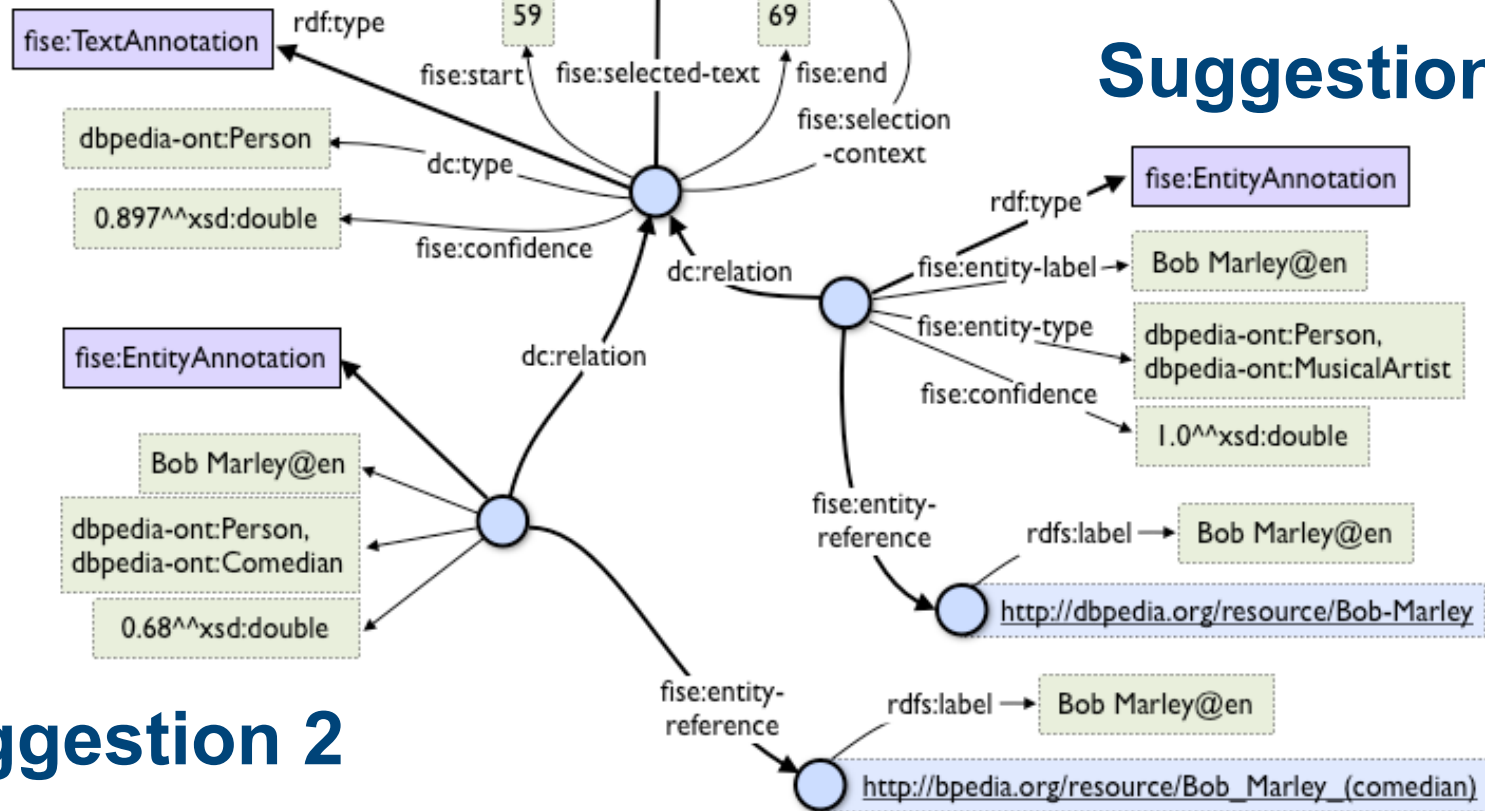
```
{  
  "@subject": "urn:enhancement-784296de-6aee-95a8-8f84-839a1e24d1b9",  
  "@type": [  
    "enhancer:Enhancement",  
    "enhancer:EntityAnnotation"  
  ],  
  "dc:created": "2012-04-13T13:43:56.016Z",  
  "dc:creator": "org.apache.stanbol.enhancer.engines.entitytagging.impl.NamedEntityTaggingEngine",  
  "dc:relation": "urn:enhancement-929e0dc8-6c5e-e44c-4c1d-c669f96d00d7",  
  "enhancer:confidence": 17396.67,  
  "enhancer:entity-label": {  
    "@literal": "Bob Marley",  
    "@language": "en"  
  },  
  "enhancer:entity-reference": "http://dbpedia.org/resource/Bob_Marley",  
  "enhancer:entity-type": [  
    "dbp-ont:MusicalArtist",  
    "foaf:Person",  
    "dbp-ont:Artist",  
    "dbp-ont:Person",  
    "owl:Thing"  
  ],  
  "enhancer:extracted-from": "urn:content-item-sha1-4186ce0dd89b27663a8ea60fc7acebceefa20174"  
},
```

**Results
as RDF**

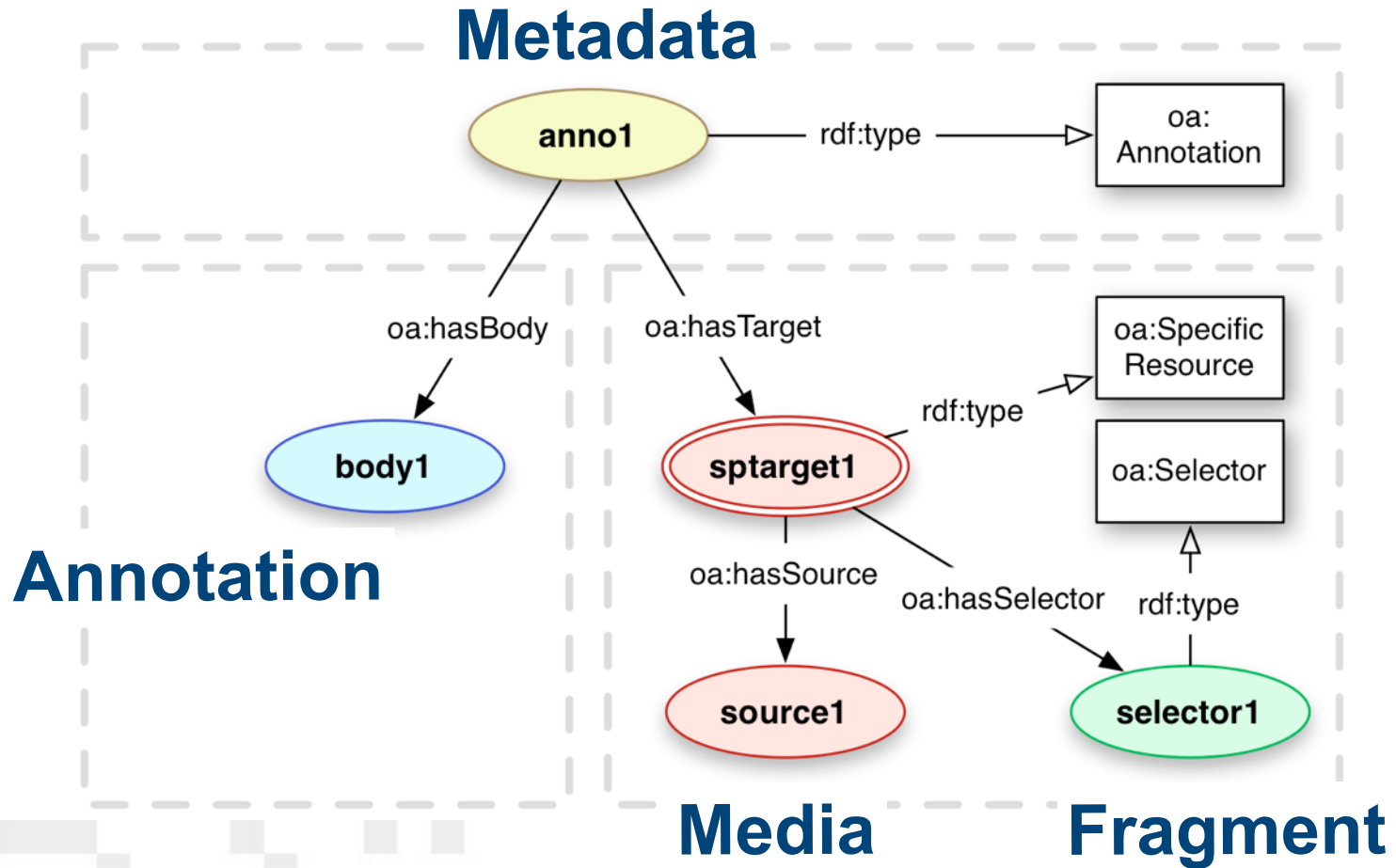
Stanbol Enhancement Structure

Mention

Apache Stanbol can detect famous entities such as Paris or **Bob Marley!**

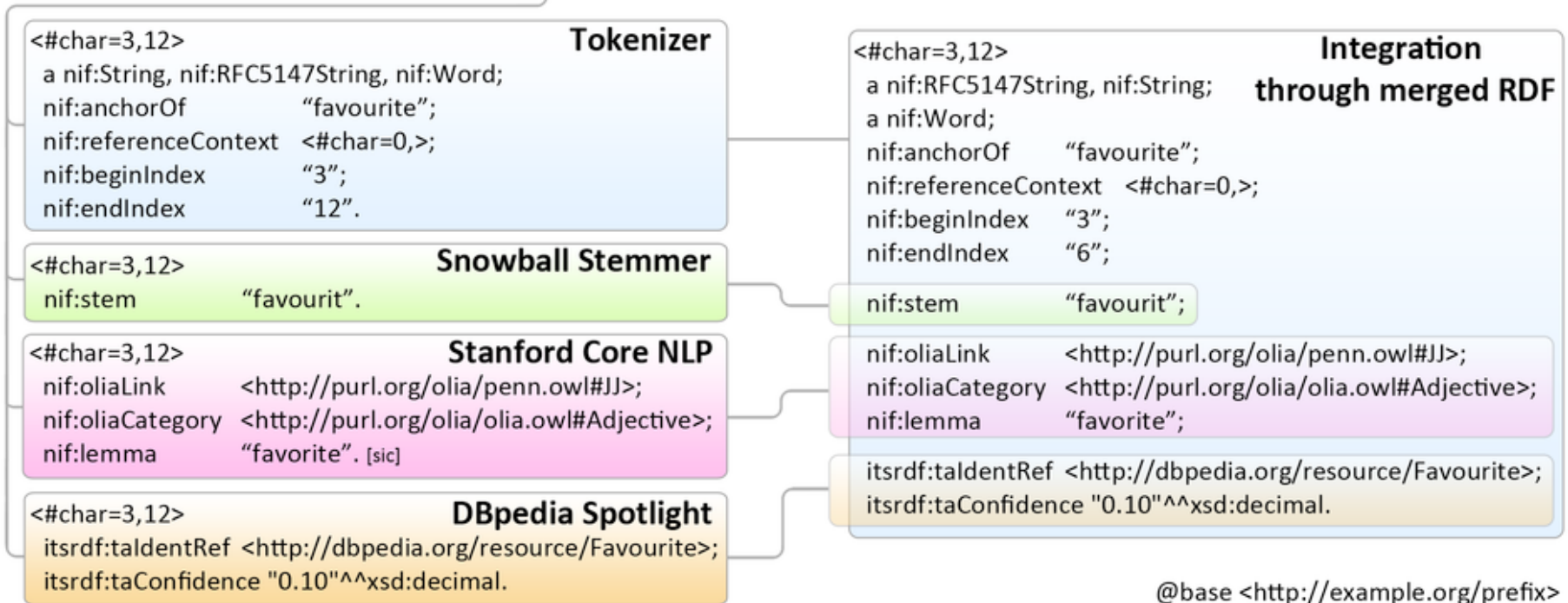


Open Annotation



NLP Interchange Format (NIF)

“My favourite actress is Natalie Portman.”



Everything



NIF Core Facts

- URI Scheme to generate Media Fragment URI's
 - <http://www.example.org/expl.txt#char=3,12>
 - allows to automatically integrate information from different Components
- Efficient Annotation Scheme
 - even suitable for word level annotations
 - selections can be encoded in the URI
 - reasoning can be used to reduce triple count
- OLiA - Ontologies of Linguistic Annotation
 - supports 34 Annotation Models and 69 Languages



Fusepool Annotation Model (1/2)

Combines

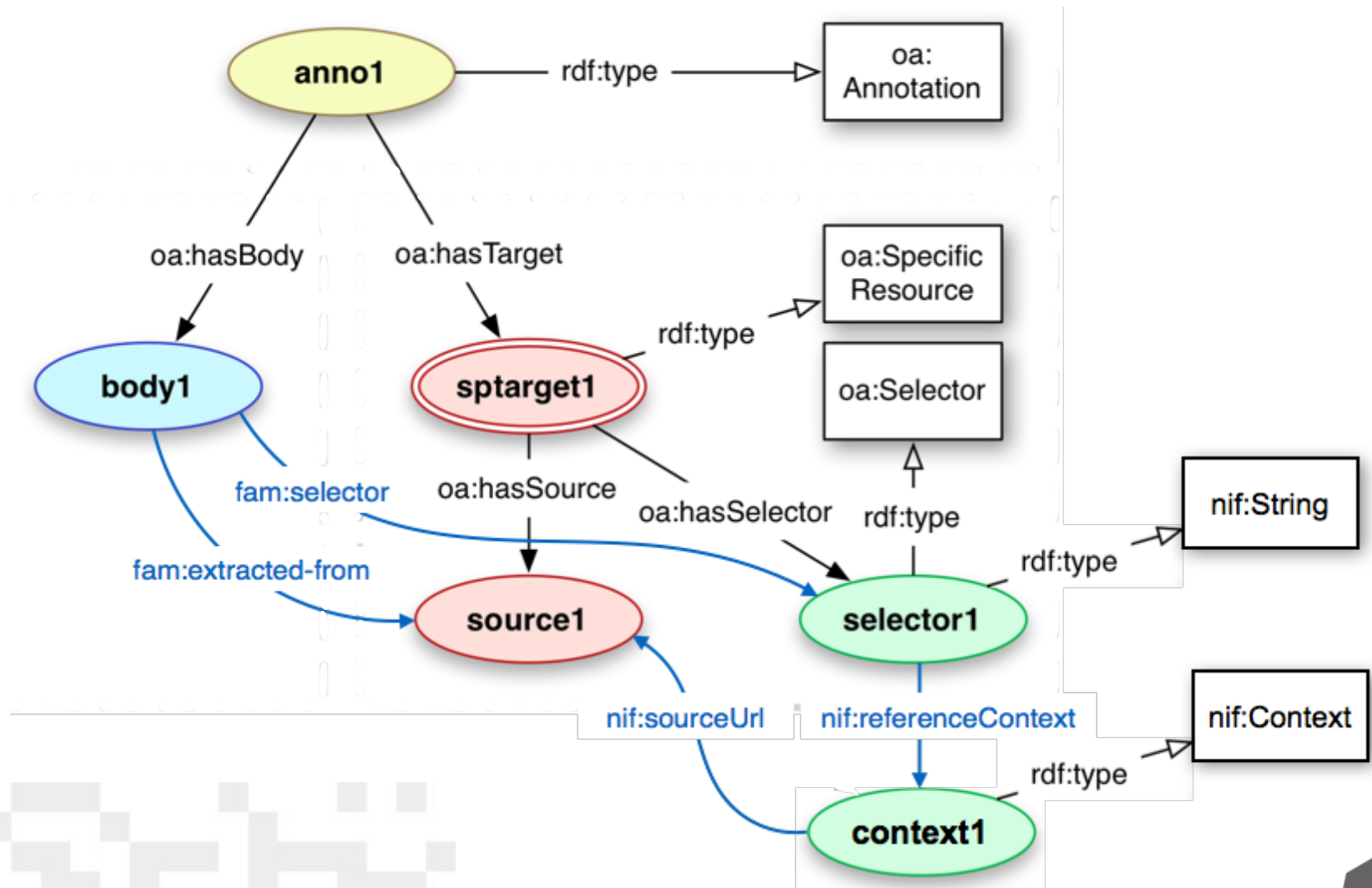
- Open Annotation ... as core annotation structure
- NIF ... to represent lower level NLP results (optional)

Extended with

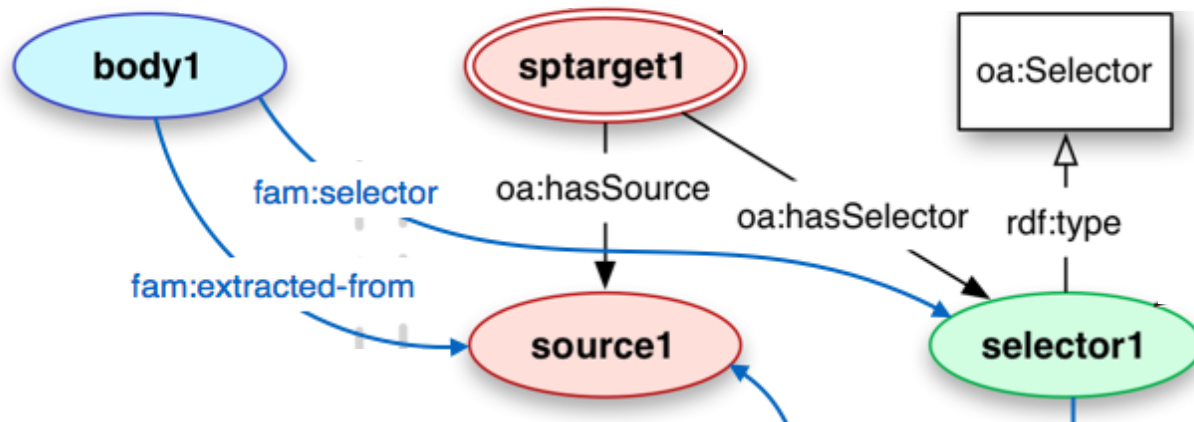
- Stanbol Enhancement Structure inspired Annotation Bodies ... for high level annotations
- Shortcuts for Media centric Annotation processing



Fusepool Annotation Model (2/2)



Media Centric Annotation Processing



```
PREFIX oa: <http://www.w3.org/ns/oa#>
```

```
PREFIX fam: <http://vocab.fusepool.info/fam#>
```

```
SELECT ?body ?source ?selector
```

```
WHERE {
```

```
  ?body a {annotation-type} ;
```

```
  fam:extracted-from ?source ;
```

```
  fam:selector ?selector .
```

```
}
```



Language Annotation

- Annotates the language of the Content

```
@prefix ex: <urn:fam-example:> .  
@prefix oa: <http://www.w3.org/ns/oa#> .  
@prefix fam: <http://vocab.fusepool.info/fam#> .  
@prefix nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#> .
```

```
ex:lang-anno-1 a fam:LanguageAnnotation ;  
    dct:language "en" ;  
    fam:confidence "0.9998"^^xsd:double ;
```



Entity Mention Annotation

- Annotates Named Entities mentioned in the Text
 - e.g from Named Entity Recognition (NER) Tools

```
ex:ent-ment-anno-1 a fam:EntityMention ;
  fam:entity-type dbo:Place;
  fam:entity-mention "Salzburg"@en
  fam:confidence "0.876"^^xsd:double ;
  fam:selector <http://www.example.com/example.txt#char=20,27> ;
  fam:extracted-from <http://www.example.com/example.txt> .
```

```
<http://www.example.com/example.txt#char=20,27>
  a fam:NifSelector, nif:String ;
  nif:referenceContext
    <http://www.example.com/example.txt#char=0>
  nif:beginIndex "20"^^xsd:int ;
  nif:endIndex "27"^^xsd:int .
```



Entity Annotation

- Annotates an Entity related to the Text
 - Entities do have an URI and are managed by Vocabularies

```
ex:keyword-anno-1 a fam:EntityAnnotation ;  
  fam:entity-reference dbr:Wolfgang_Amadeus_Mozart ;  
  fam:entity-type dbo:Person;  
  fam:entity-label "Wolfgang Amadeus Mozart"@en ;  
  fam:confidence "0.789"^^xsd:double ;  
  fam:extracted-from <http://www.example.com/example.txt> .
```

- Entity Annotations do not define the mention(s) of the Entity in the Text.



Linked Entity Annotation

- Combines an Entity Mention with a Linked Entity
 - Links an mention in the Text with an Entity as defined by a Vocabulary.

```
ex:linked-entity-anno-1
  a fam:LinkedEntity, fam:EntityMention, fam:EntityAnnotation ;
  fam:entity-reference dbr:Salzburg ;
  fam:entity-type dbo:Place;
  fam:entity-mention "Salzburg"@en ;
  fam:entity-label "Salzburg"@en ;
  fam:confidence "0.893"^^xsd:double ;
  fam:selector <http://www.example.com/example.txt#char=20,27> ;
  fam:extracted-from <http://www.example.com/example.txt> .
```



Entity Suggestion

■ Suggest multiple Entities for a Mention

```
ex:entity-linking-choice-anno-1 a fam:EntityLinkingChoice ;
  fam:entity-mention "Salzburg"@en ;
  oa:item ex:entity-suggestion-1, ex:entity-suggestion-2 .
  fam:selector <http://www.example.com/example.txt#char=20,27> ;
  fam:extracted-from <http://www.example.com/example.txt> .
```

```
ex:entity-suggestion-1 a fam:EntitySuggestion;
  fam:entity-reference dbr:Salzburg ;
  fam:entity-label "Salzburg"@en ;
  fam:entity-type dbo:Place ;
  fam:confidence "0.973"^^xsd:double ;
  fam:extracted-from <http://www.example.com/example.txt> .
```

```
ex:entity-suggestion-2 a fam:EntitySuggestion;
  fam:entity-reference dbr:Salzburg_(state) ;
  fam:entity-label "Salzburg"@en ;
  fam:entity-type dbo:Place ;
  fam:confidence "0.573"^^xsd:double ;
  fam:extracted-from <http://www.example.com/example.txt> .
```



Topic Classification

- Classifies a Content along multiple Categories

```
ex:topic-classification-anno-1 a fam:TopicClassification ;  
  fam:classification-scheme my:ConceptScheme ;  
  oa:item ex:topic-anno-1, ex:topic-anno-2 .  
  fam:selector <http://www.example.com/example.txt#char=0> ;  
  fam:extracted-from <http://www.example.com/example.txt> .
```

```
ex:ex:topic-anno-1 a fam:TopicAnnotation;  
  fam:topic-reference my:ClassicalComposers ;  
  fam:topic-label "Classical Composers"@en ;  
  fam:confidence "0.872"^^xsd:double .  
  fam:extracted-from <http://www.example.com/example.txt> .
```

```
ex:topic-anno-2 a fam:TopicAnnotation;  
  fam:topic-reference my:Austria ;  
  fam:topic-label "Salzburg"@en ;  
  fam:confidence "0.743"^^xsd:double .  
  fam:extracted-from <http://www.example.com/example.txt> .
```



Stanbol Enhancer Support

■ NIF 2.0 Transformation Engine [1]

- part of the `org.apache.stanbol.enhancer.engines.nlp2rdf` module
- version: `>= 0.12.1` and `1.0.0-SNAPSHOT`
- serializes the Analyzed Text Content Part as NIF 2.0

■ FISE to FAM Converter Engine [2]

- provided by the `eu.fusepool.p3.stanbol-engines-fise2fam:stanbol-engines-fise2fam` module
- version: `1.0.0`
- converts the RDF of the Stanbol Enhancement Structure to the FAM

[1] <http://stanbol.apache.org/docs/trunk/components/enhancer/engines/nif20>

[2] <https://github.com/fusepoolP3/p3-stanbol-engine-fam>



Demo Setup (1/2)

- Analysis Chain configuration

-  for NLP Annotations

- DBpedia Linking using  **MachineLinking** [1]

- NIF 2.0 Engine

- Text Annotation New Model Engine

- for prefix/suffix information of Selectors

- FISE 2 FAM Engine

apachecon-demo chain

⚙ **langdetect** (required , LanguageDetectionE

⚙ **opennlp-sentence** (required , OpenNlpS

⚙ **opennlp-token** (required , OpenNlpToken

⚙ **opennlp-pos** (required , OpenNlpPosTagg

⚙ **machinelinkingAnnotate** (required , M

⚙ **text-annotation-new-model** (requirec

⚙ **fise2fam** (required , Fise2FamEngine)

[1] <https://github.com/michelemostarda/machinelinking-istanbol-enhancement-engine>



Demo Setup (2/2)

- Query Enhancement Results

-  as RDF Triple Store
- **Marmotta** and SPARQL Endpoint

- Squebi as SPARQL editor [1]

- Demo Data

- 6 English, 4 German, 4 Italian, 4 French and 4 Spanish news articles about Ebola

[1] <https://github.com/tkurz/squebi>



Demo



Stanbol Enhancer Analysis

[Web View](#) [REST API](#)

Apache Stanbol Enhancer

Enhancement Chain: **apachecon-demo** all 7 engines available

[< List of Enhancement Chains >](#)

Paste some text below and submit the form to let the Enhancement Chain apachecon-demo enhance it:

```
Texas healthcare workers at risk of Ebola asked to stay out of public
Seventy-five staff members of Dallas hospital asked to sign 'binding legal order' that states they will avoid public spaces
Healthcare workers deemed to be at risk of contracting Ebola after dealing with a patient who died from the virus in Texas are being asked
to sign voluntary agreements to stay away from the public, after Dallas authorities decided against declaring a state of emergency.
Seventy-five staff members from Texas Health Presbyterian hospital are being given a "binding legal document and order" that states they
will avoid public transport, not go to areas where large numbers of people congregate, and continue to be monitored twice a day for
symptoms, county judge Clay Jenkins said on Thursday.
Any of those involved in the care of Thomas Eric Duncan who refuse to sign the agreement would be subject to a legal control order,
Jenkins told reporters after a meeting of the county commissioners court in downtown Dallas. "All the remedies of the law are available,"
he said. However he said he believed this would not be necessary. "These are hometown healthcare heroes," he said. "They're not going to
jail."
```

Output format:

Enhancement Process Metadata

Execution of Chain **apachecon-demo** completed in 735ms.

Raw RDF output

```
<urn:enhancement-51fce9e1-6d53-982b-a115-44ca50ede735-annotation>
  a      <http://www.w3.org/ns/oa#Annotation> ;
  <http://www.w3.org/ns/oa#annotatedAt>
    "2014-11-05T09:35:20.366Z"^^<http://www.w3.org/2001/XMLSchema#dateTime> ;
  <http://www.w3.org/ns/oa#annotatedBy>
    "com.machinelinking.stanbol.enhancer.engines.machinelinking.impl.MLAnnotateEnhancementEngine"^^<http://www.w3.org
/2001/XMLSchema#string> ;
  <http://www.w3.org/ns/oa#hasBody>
    <urn:enhancement-51fce9e1-6d53-982b-a115-44ca50ede735> ;
  <http://www.w3.org/ns/oa#hasTarget>
    <urn:enhancement-51fce9e1-6d53-982b-a115-44ca50ede735-sptarget> ;
```



Entity Mention Result (Example)

property	hasValue	isValueOf
http://vocab.fusepool.info/fam#entity-mention	WHO @en	
http://vocab.fusepool.info/fam#entity-type	http://dbpedia.org/ontology/Agent	
http://vocab.fusepool.info/fam#extracted-from	http://localhost:8080/apachecon-demo/data/news6.txt	
http://vocab.fusepool.info/fam#selector	http://localhost:8080/apachecon-demo/data/news6.txt#char=751,754	
rdf:type	http://vocab.fusepool.info/fam#EntityLinkingChoice	
rdf:type	http://vocab.fusepool.info/fam#EntityMention	
rdf:type	http://www.w3.org/ns/oa#Choice	
http://www.w3.org/ns/oa#hasBody		urn:enhancement-01895b4c-900d-574a-6828-24d8dc8a934f-annotation
http://www.w3.org/ns/oa#item	urn:enhancement-3c7a01ac-dcba-422d-09ff-440e5eb22d58	



Selector Result (Example)

property	hasValue	isValueOf
http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#after) can decl @en	
http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#anchorOf	WHO @en	
http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#before	nization (@en	
http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#beginIndex	751	
http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#endIndex	754	
http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#referenceContext	http://localhost:8080/apachecon-demo/data/news6.txt#char=0	



Topic Annotation (Example)

property	hasValue	isValueOf
http://vocab.fusepool.info/fam#confidence	0.8540145754814148	
http://vocab.fusepool.info/fam#extracted-from	http://localhost:8080/apachecon-demo/data/news2.txt	
http://vocab.fusepool.info/fam#topic-label	Medicine	
http://vocab.fusepool.info/fam#topic-reference	http://dbpedia.org/resource/Medicine	
rdf:type	http://vocab.fusepool.info/fam#TopicAnnotation	
http://www.w3.org/ns/oa#hasBody		urn:enhancement-736cf055-e843-4b50-3586-2ad2a32e8fcd-annotation
http://www.w3.org/ns/oa#item		urn:enhancement-9fe634ea-8567-efff-bad0-5967d6d98abc

Query Mentioned Entities

```
PREFIX nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#>  
PREFIX oa: <http://www.w3.org/ns/oa#>  
PREFIX fam: <http://vocab.fusepool.info/fam#>
```

```
SELECT DISTINCT ?doc ?mention ?start ?end ?entity WHERE {  
  ?mention a <http://vocab.fusepool.info/fam#EntityMention> ;  
    fam:extracted-from ?doc ;  
    fam:entity-mention ?mention ;  
    fam:selector ?selector ;  
    oa:item ?suggestion .  
  ?selector nif:beginIndex ?start ;  
    nif:endIndex ?end .  
  ?suggestion fam:entity-reference ?entity .  
}  
ORDER BY ?doc ASC(xsd:integer(?start))  
LIMIT 100
```

<http://localhost:8080/apachecon-demo/data/news-de-1.txt>

EU @de

223

225

http://dbpedia.org/resource/European_Union



Query Topic Annotations

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#>
PREFIX oa: <http://www.w3.org/ns/oa#>
PREFIX fam: <http://vocab.fusepool.info/fam#>

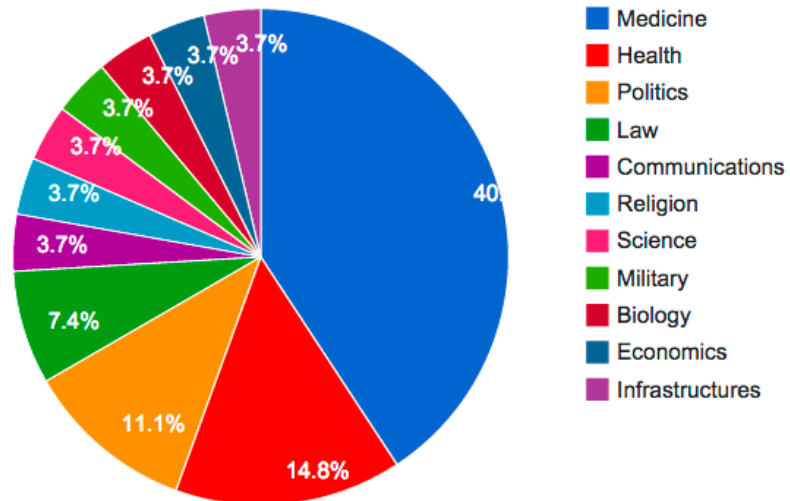
SELECT DISTINCT ?confidence ?tag ?topic WHERE {
  ?m a <http://vocab.fusepool.info/fam#TopicAnnotation> ;
    fam:extracted-from <http://localhost:8080/apachecon-demo/data/news5.txt> ;
    fam:confidence ?confidence ;
    fam:topic-reference ?topic ;
    fam:topic-label ?tag .
} ORDER BY DESC(xsd:double(?confidence))
LIMIT 100
```

confidence	tag	topic
0.7198156118392944	Medicine	http://dbpedia.org/resource/Medicine
0.14788953959941864	Health	http://dbpedia.org/resource/Health
0.13229481875896454	Statistics	http://dbpedia.org/resource/Statistics

Categories Overview

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#>
PREFIX oa: <http://www.w3.org/ns/oa#>
PREFIX fam: <http://vocab.fusepool.info/fam#>
```

```
SELECT DISTINCT ?tag (COUNT (?tag) AS ?count) WHERE {
  ?m a <http://vocab.fusepool.info/fam#TopicAnnotation> ;
  fam:extracted-from ?doc ;
  fam:confidence ?confidence ;
  fam:topic-label ?tag .
  FILTER ( xsd:float(?confidence) >= "0.33"^^xsd:double ) .
} GROUP BY ?tag
ORDER BY DESC(?count)
```





salzburgresearch

Rupert Westenthaler

Researcher

Salzburg Research Forschungsgesellschaft mbH

Jakob Haringer Straße 5/3 | 5020 Salzburg, Austria

T +43.662.2288-413 | F -222

rupert.westenthaler@salzburgresearch.at

Fusepool **p3** 

<http://p3.fusepool.eu/>

