



Seminar: Annotations and Natural Language Processing

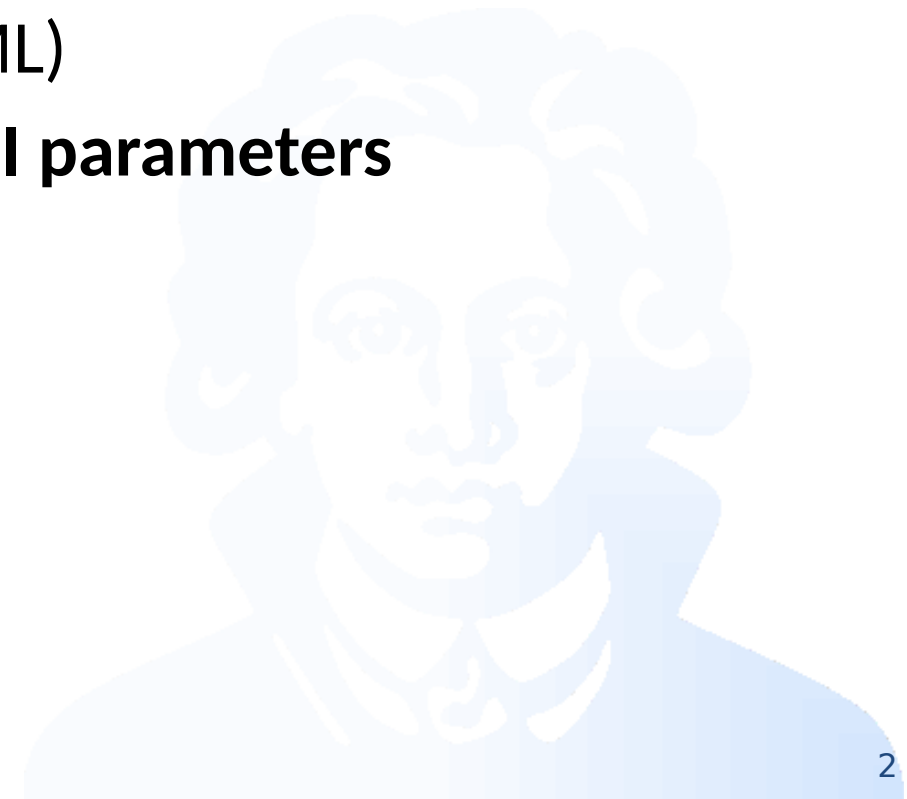
Christian Chiarcos
Goethe University Frankfurt, Germany
University of Cologne, Germany



Motivation

The “Developers Nightmare”

- Many NLP tools fulfill similar functions but are **not interoperable**
- **Heterogeneous** output formats (JSON, XML)
- NLP Web services with **heterogeneous API parameters**
- **Heterogeneous ways of annotating text**





Annotations and NLP

- WebAnnotation / Open Annotation
- NLP Interchange Format
- CoNLL-RDF
- Hands-on sessions (annotation/NLP track):
 - Today: Transformation with Fintan (Max Ionov, Christian Fäth)
 - Tomorrow: Corpora (Francesco Mambrini, Max Ionov)
 - Thursday: Linking with Dictionaries (Max Ionov, Christian C., Fahad Khan)
 - Friday: NLP Workflows with Teanga (Bernardo Stearns)



Web Annotation / Open Annotation

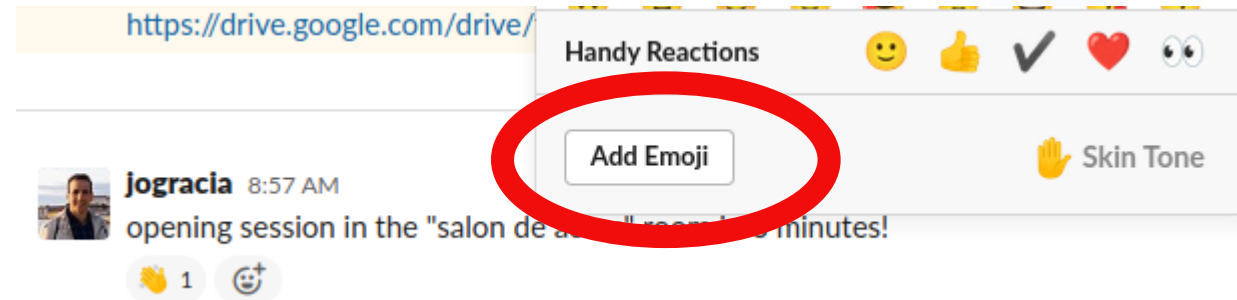
- Problem addressed
 - A lot of content is on the web
 - Images, videos, text, ...
 - And being commented on (tags, etc.)
 - Tools for annotation / tagging





Web Annotation / Open Annotation

- Problem addressed
 - A lot of content is on the web
 - Images, videos, text, ...
 - And being commented on (tags, etc.)
 - Tools for annotation / tagging



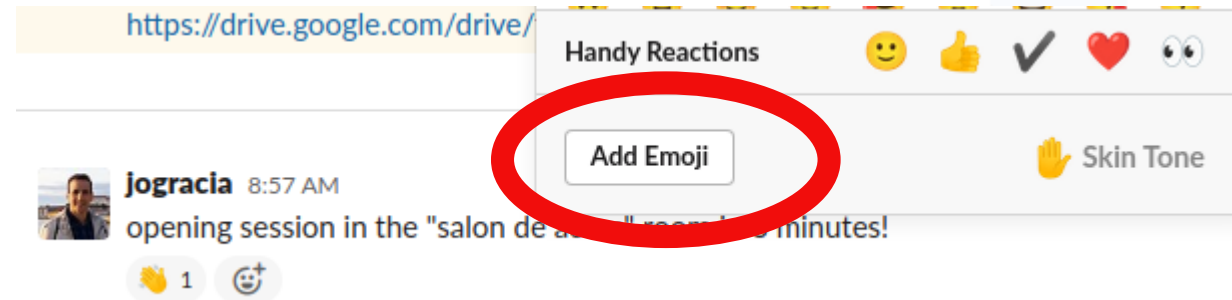


Web Annotation / Open Annotation

- Problem addressed
 - A lot of content is on the web
 - Images, videos, text, ...
 - And being commented on (tags, etc.)
 - Tools for annotation / tagging

over any type of static content available on the web

Standoff annotation with JSON-LD





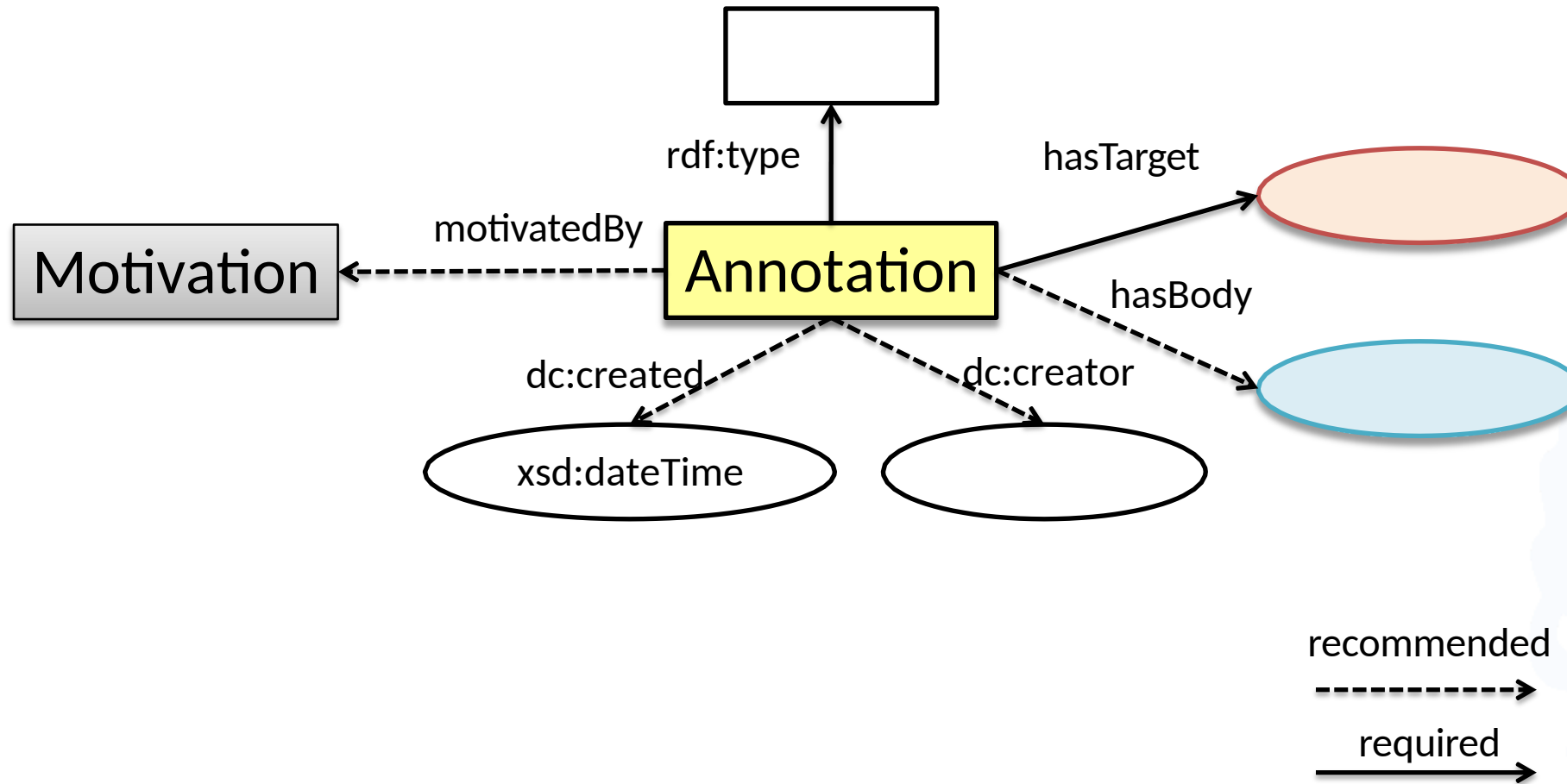
Web Annotation / Open Annotation

- W3C Open Annotation Community Group
 - <https://www.w3.org/community/openannotation/> (2012-2014)
 - mostly driven by bioinformatics, but generic formalism for annotating web content
- Web Annotation (W3C recommendations, Feb 2017)
 - Data Model: <https://www.w3.org/TR/annotation-model>
 - general description
 - Vocabulary: <https://www.w3.org/TR/annotation-vocab>
 - ontology
 - Protocol: <https://www.w3.org/TR/annotation-protocol>
 - retrieving and manipulating annotations
 - serialization: *must* JSON-LD, *should* Turtle, *may* provide other RDF serializations



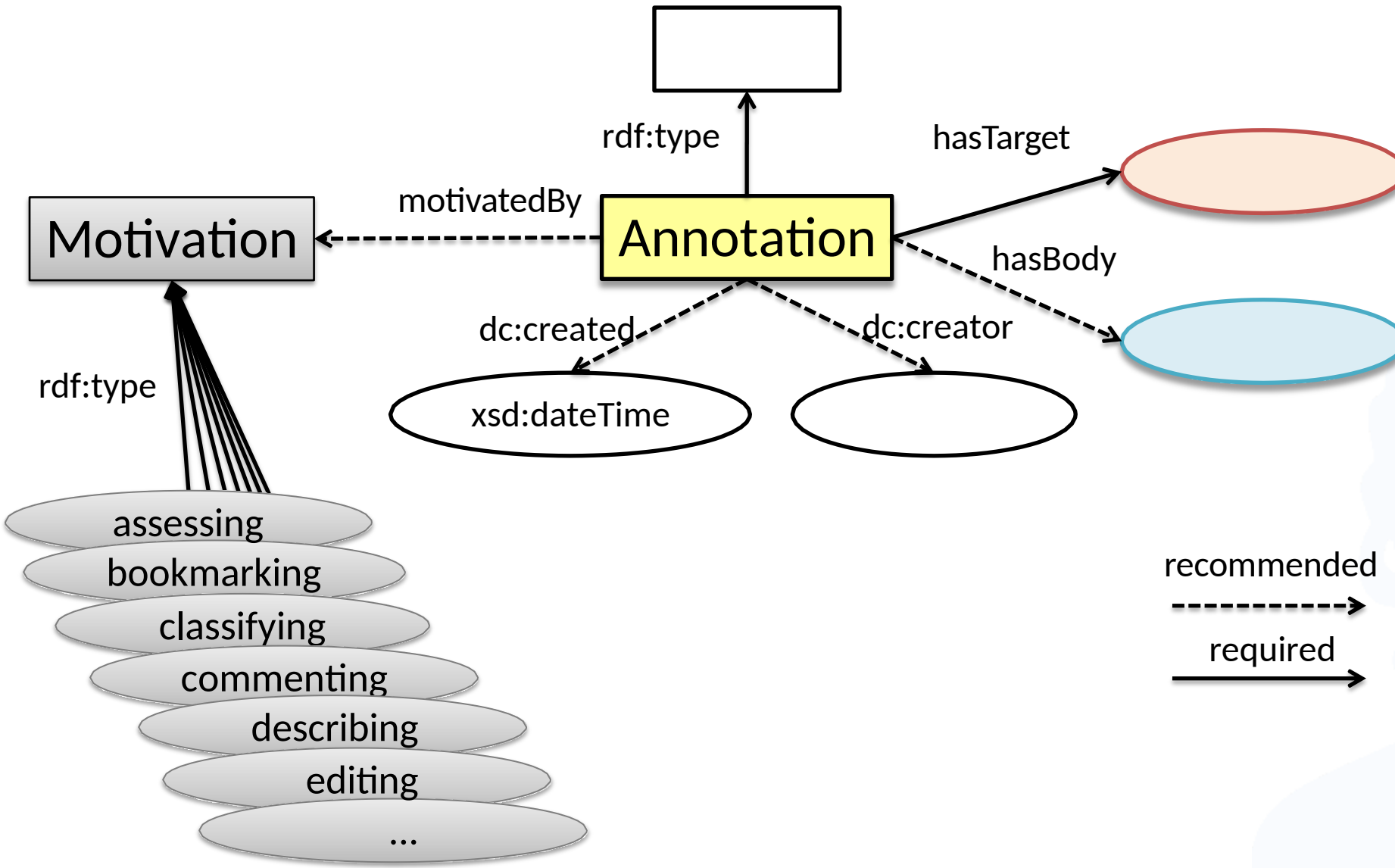
Web Annotation: Annotation

<https://www.w3.org/TR/annotation-vocab/>



Web Annotation: Annotation

<https://www.w3.org/TR/annotation-vocab/>





Web Annotation: Target and Body

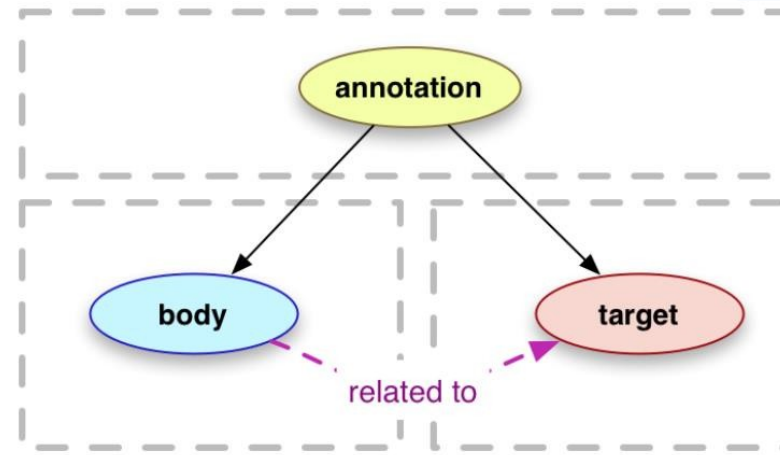
<https://www.w3.org/TR/annotation-model/>

■ body

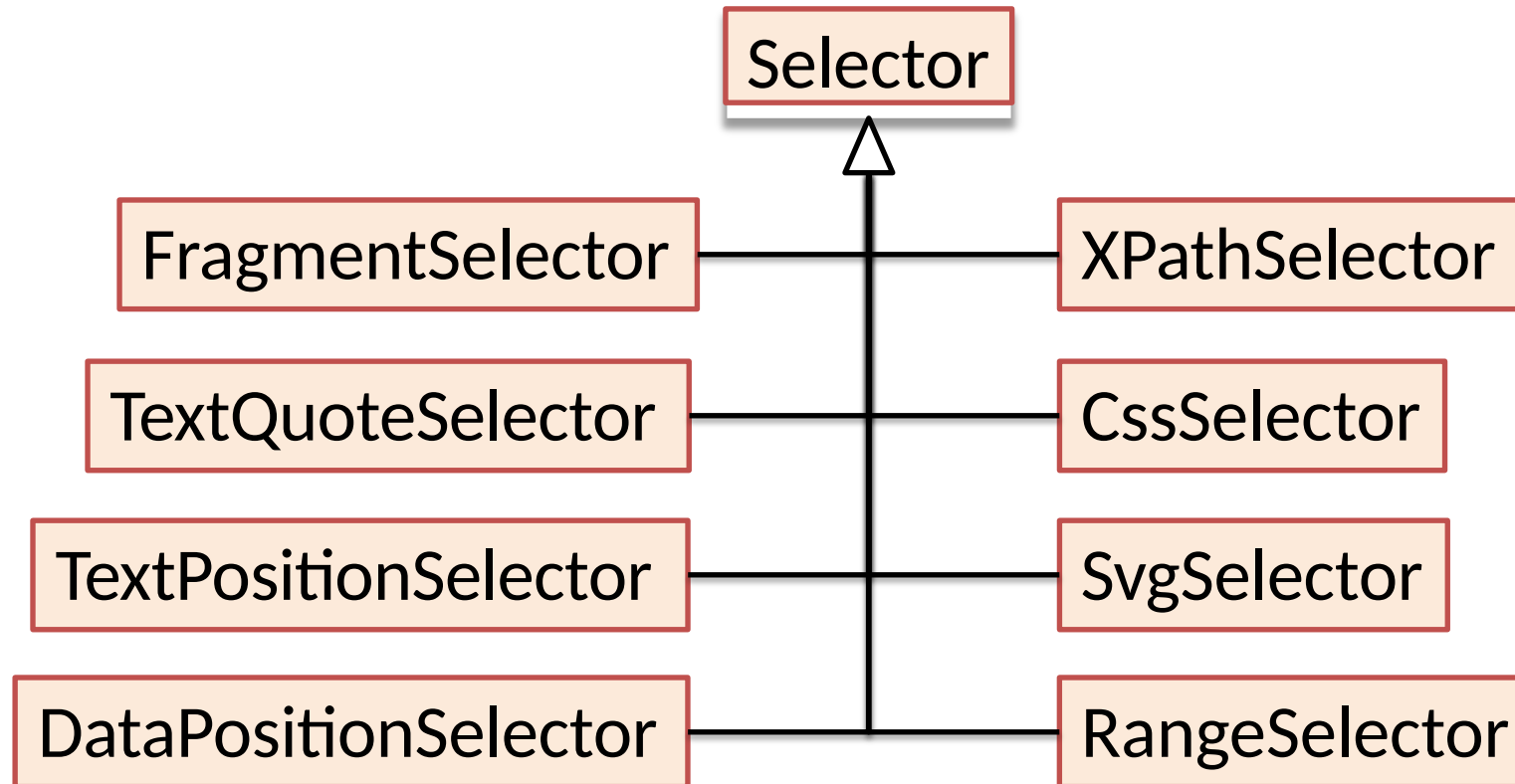
- ❑ element containing the annotation
- ❑ object property: *oa:hasBody* (any RDF object)
- ❑ datatype property: *oa:bodyValue* (strings)

■ target

- ❑ element being annotated
- ❑ any RDF object, *including*
 - *oa:Selector* (more in a second)



oa:Selector – e.g. possible targets



or: just any URI ;)



Named Entity Annotations

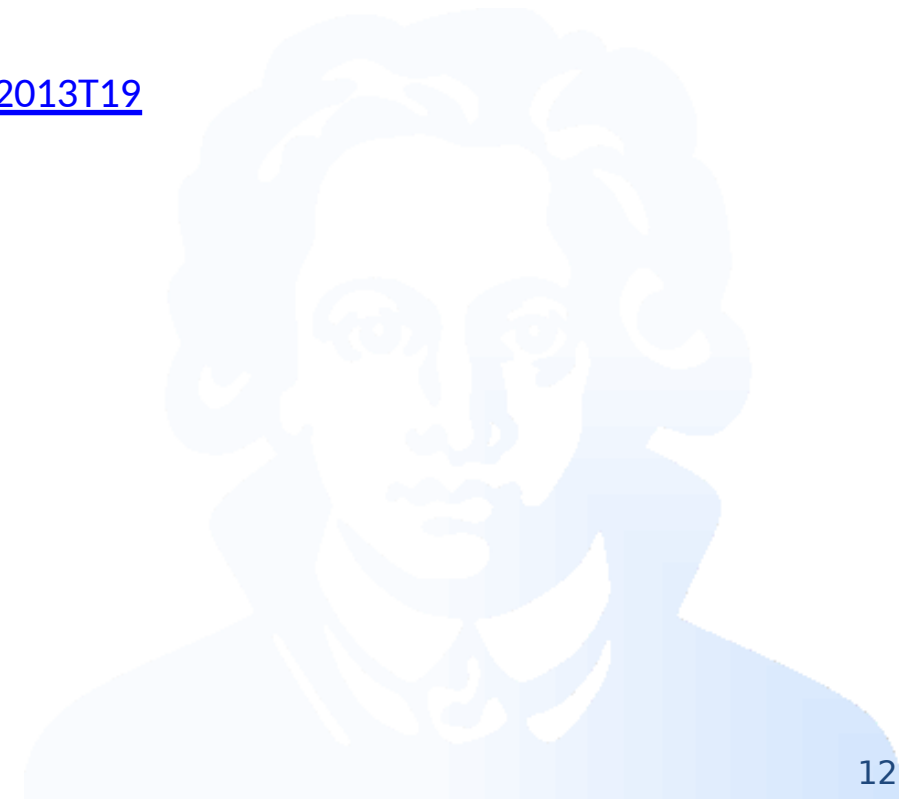
Secretary of State **James Baker**, who accompanied President **Bush** to **Costa Rica**, told reporters **Friday**: “I have no reason to deny reports that some **Contras** ambushed some **Sandinista** soldiers. ”

organizations, persons, geopolitical entities, dates,

James	B-PERSON
Baker	E-PERSON
told	O
reporters	O
Friday	S-DATE
:	O

OntoNotes corpus, wsj-0655

<https://catalog.ldc.upenn.edu/LDC2013T19>





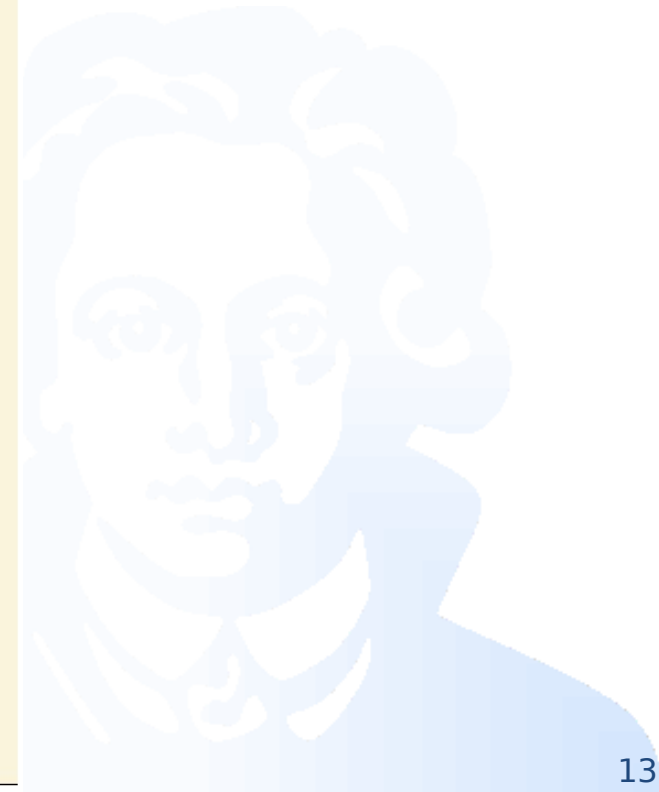
Named Entity Annotations (JSON-LD)

Secretary of State James Baker, who accompanied President Bush to Costa Rica, told reporters Friday: “I have no reason to deny reports that some Contras ambushed some Sandinista soldiers.”

James	B-PERSON
Baker	E-PERSON
told	O
reporters	O
Friday	S-DATE
:	O

```

1  {
2    "@graph": [
3      {
4        "@context": "http://www.w3.org/ns/anno.jsonld",
5        "id": "http://example.org/enamex2",
6        "type": [
7          "Annotation",
8          "https://catalog.ldc.upenn.edu/docs/LDC2007T21/
          ontonotes-1.0-documentation.pdf#ENAMEX"
9        ],
10       "body": {
11         "type" : "TextualBody",
12         "value" : "PERSON",
13         "format" : "text/plain"
14       },
15       "target": {
16         "source": "https://catalog.ldc.upenn.edu/
          ldc2013t19/data/files/data/english/
          annotations/nw/wsj/06/wsj_0655.name",
17         "selector": {
18           "type": "TextQuoteSelector",
19           "exact": "James Baker"
20         }
21       }
22     ]
  }
```





Named Entity Annotations (Turtle)

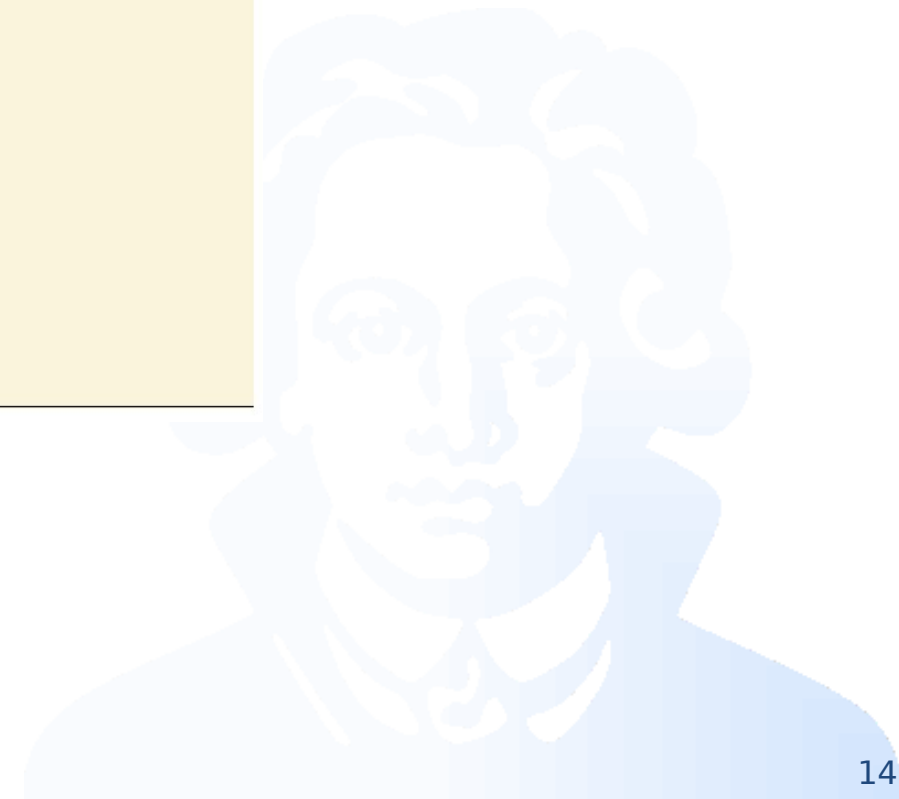
Secretary of State James Baker, who accompanied President Bush to Costa Rica, told reporters Friday: “I have no reason to deny reports that some Contras ambushed some Sandinista soldiers. ”

James	B-PERSON
Baker	E-PERSON
told	O
reporters	O
Friday	S-DATE
:	O

```

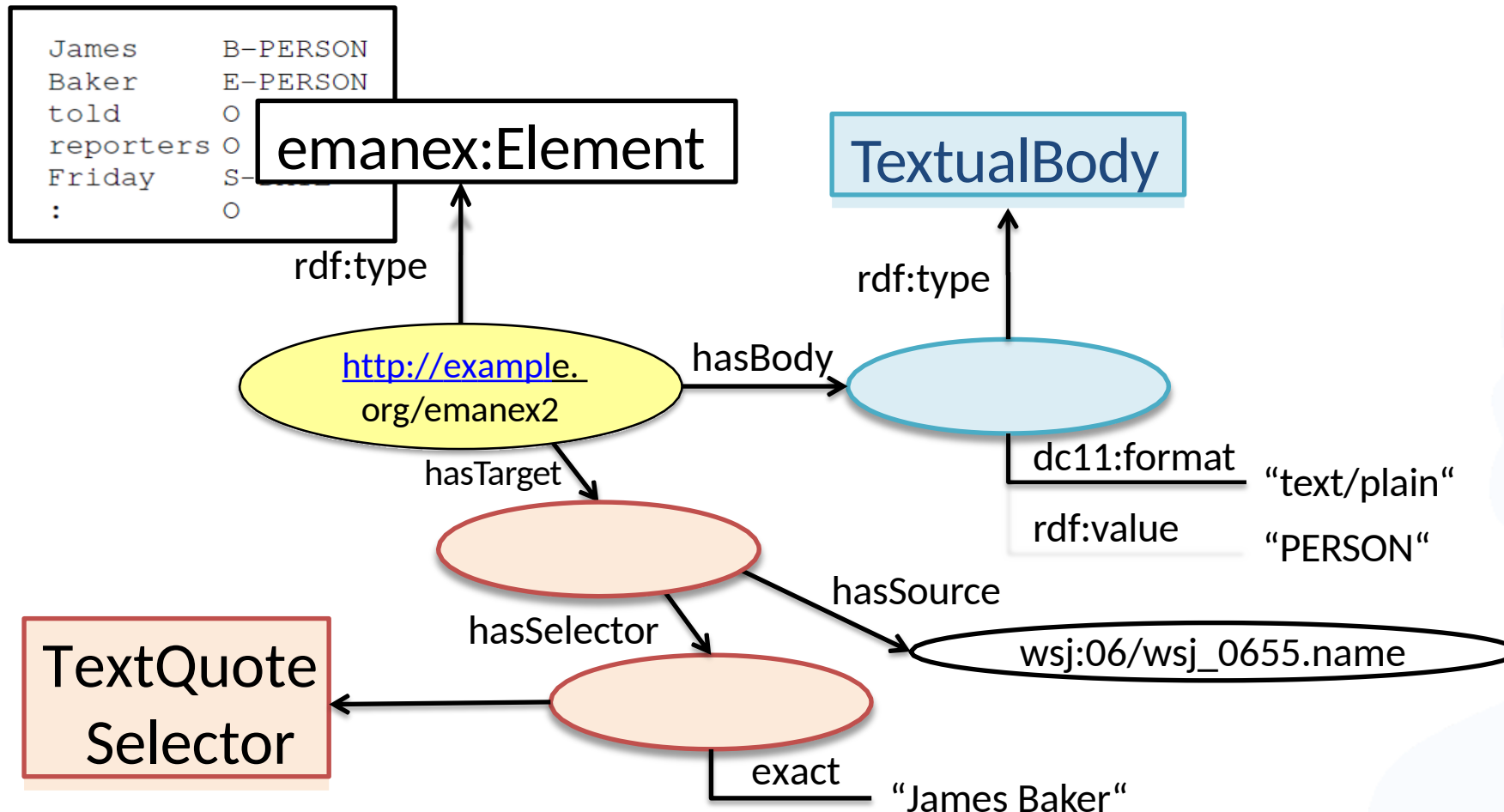
1 <http://example.org/enamex2>
2   a oa:Annotation, on:ENAMEX ;
3   oa:hasBody [
4     a oa:TextualBody ;
5     dc11:format "text/plain"^^xsd:string ;
6     rdf:value "PERSON"^^xsd:string
7   ] ;
8   oa:hasTarget [
9     oa:hasSelector [
10      a oa:TextQuoteSelector ;
11      oa:exact "James Baker"^^xsd:string
12    ] ;
13    oa:hasSource wsj:06/wsj_0655.name
14  ] .

```



Named Entity Annotations

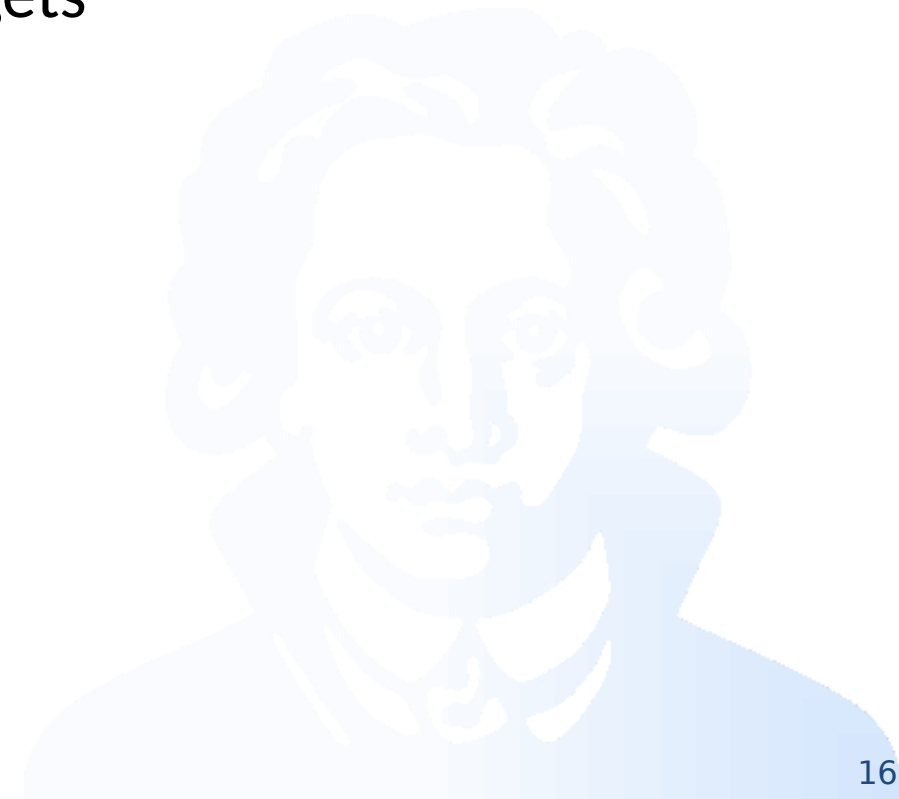
Secretary of State James Baker, who accompanied President Bush to Costa Rica, told reporters Friday: “I have no reason to deny reports that some Contras ambushed some Sandinista soldiers.”





Web Annotation: Overview

- relatively good uptake
 - esp. in bioinformatics
- reification
 - annotation as $n:m$ relation between bodies & targets
 - with metadata
- powerful
 - annotate all instances of a string at once using a *oa:TextQuoteSelector*
- very verbose
 - „X is a person according to EMANEX annotations“
takes 11 triples





Named Entity Annotations (Turtle)

Secretary of State James Baker, who accompanied President Bush to Costa Rica, told reporters Friday: “I have no reason to deny reports that some Contras ambushed some Sandinista soldiers. ”

James	B-PERSON
Baker	E-PERSON
told	O
reporters	O
Friday	S-DATE
:	O

```

1 <http://example.org/enamex2>
2   a oa:Annotation, on:ENAMEX ;
3   oa:hasBody [
4     a oa:TextualBody ;
5     dc11:format "text/plain"^^xsd:string ;
6     rdf:value "PERSON"^^xsd:string
7   ] ;
8   oa:hasTarget [
9     oa:hasSelector [
10      a oa:TextQuoteSelector ;
11      oa:exact "James Baker"^^xsd:string
12    ] ;
13    oa:hasSource wsj:06/wsj_0655.name
14  ] .

```

notational shorthand: *oa:bodyValue* for string-value bodies

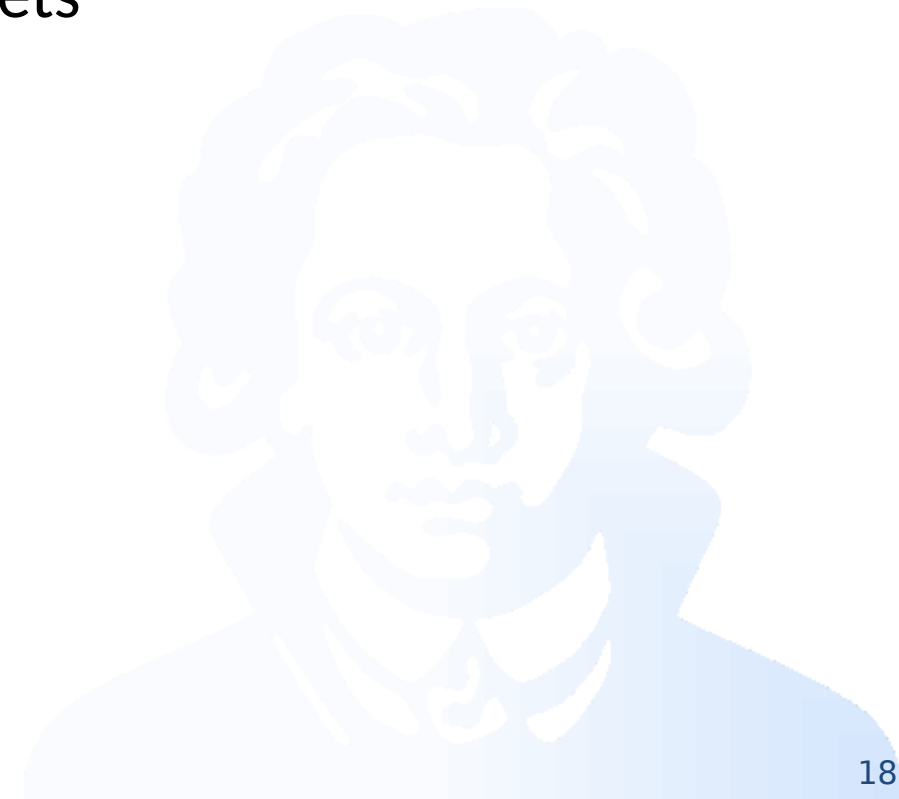
3-7 *oa:bodyValue*
"PERSON"^^xsd:string ;

=> 4 triples replaced by 1



Web Annotation: Overview

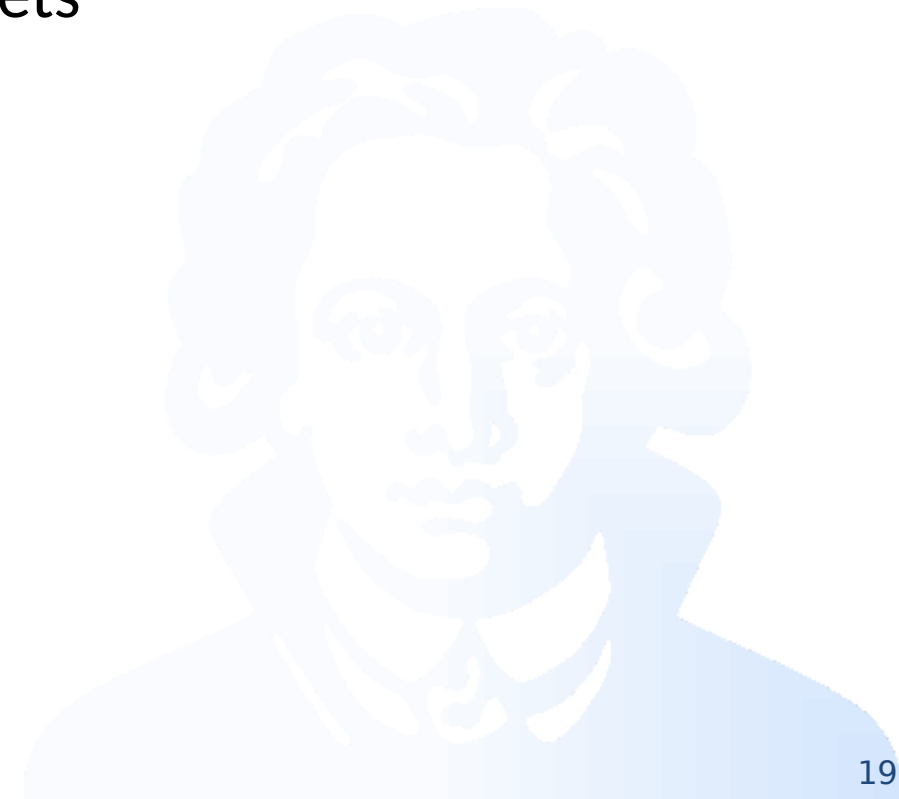
- relatively good uptake
 - esp. in bioinformatics
- reification
 - annotation as $n:m$ relation between bodies & targets
 - with metadata
- powerful
 - annotate all instances of a string at once using a *oa:TextQuoteSelector*
- rather verbose
 - „X is a person according to EMANEX annotations“
takes ± 7 triples





Web Annotation: Overview

- relatively good uptake
 - esp. in bioinformatics
- reification
 - annotation as $n:m$ relation between bodies & targets
 - with metadata
- powerful
 - annotate all instances of a string at once using a *oa:TextQuoteSelector*
- rather verbose
- no linguistic data structures





 **Nexus**
Linguarum

4th Summer Datathon on Linguistic Linked Open Data
Cercedilla, Spain
30th May – 3rd June 2022

NIF - NLP Interchange Format

NIF URI schema & NIF ontologies



NIF - NLP Interchange Format

<http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core/>

*The **NLP Interchange Format (NIF)** is an RDF/OWL-based format that aims to achieve interoperability between NLP tools, language resources and annotations.*



NIF - NLP Interchange Format

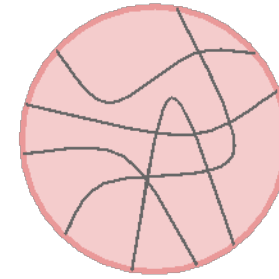
<http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core/>

- Way to **mint URIs** for arbitrary strings on the Web
- Logical **formalisation of strings** and **annotations** via an ontology
- Easy and human **understandable format**
- Builds on **existing standards** (RDF, LAF/GrAF, RFC 5147)
- **Reuses existing RDF tools** and implementations
- **Decreases development costs** for integration of tools and resources

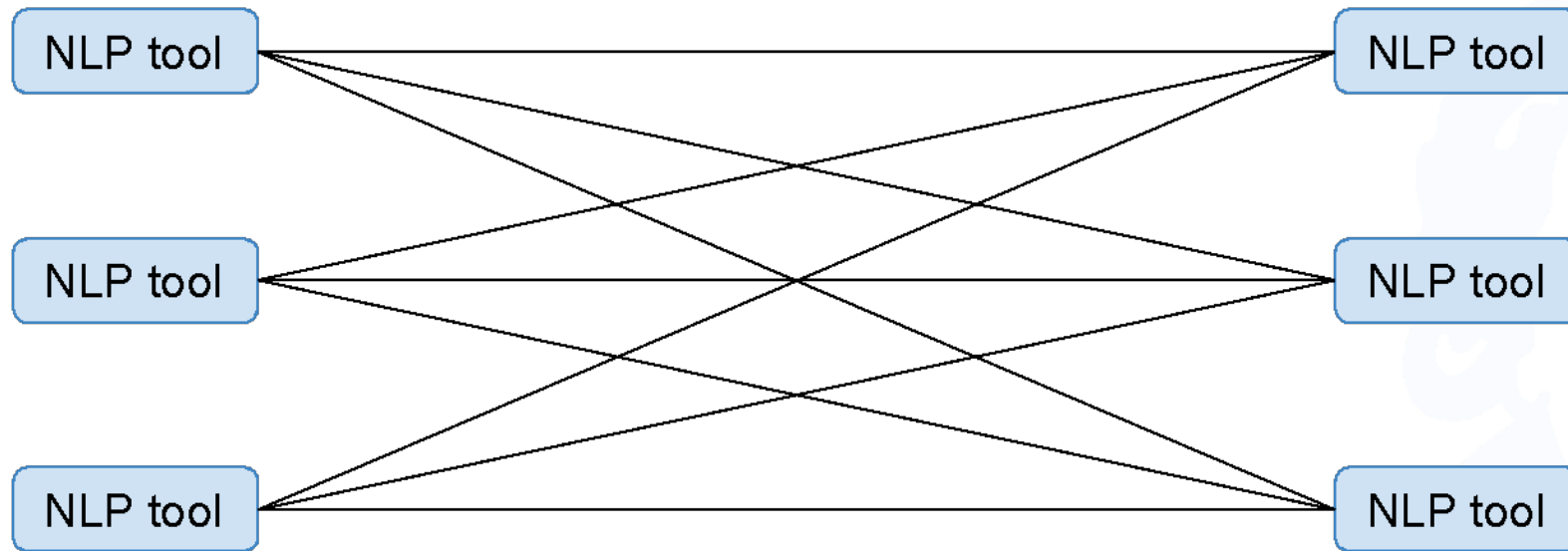
Pre-NIF Spaghetti Architecture

Need for integration

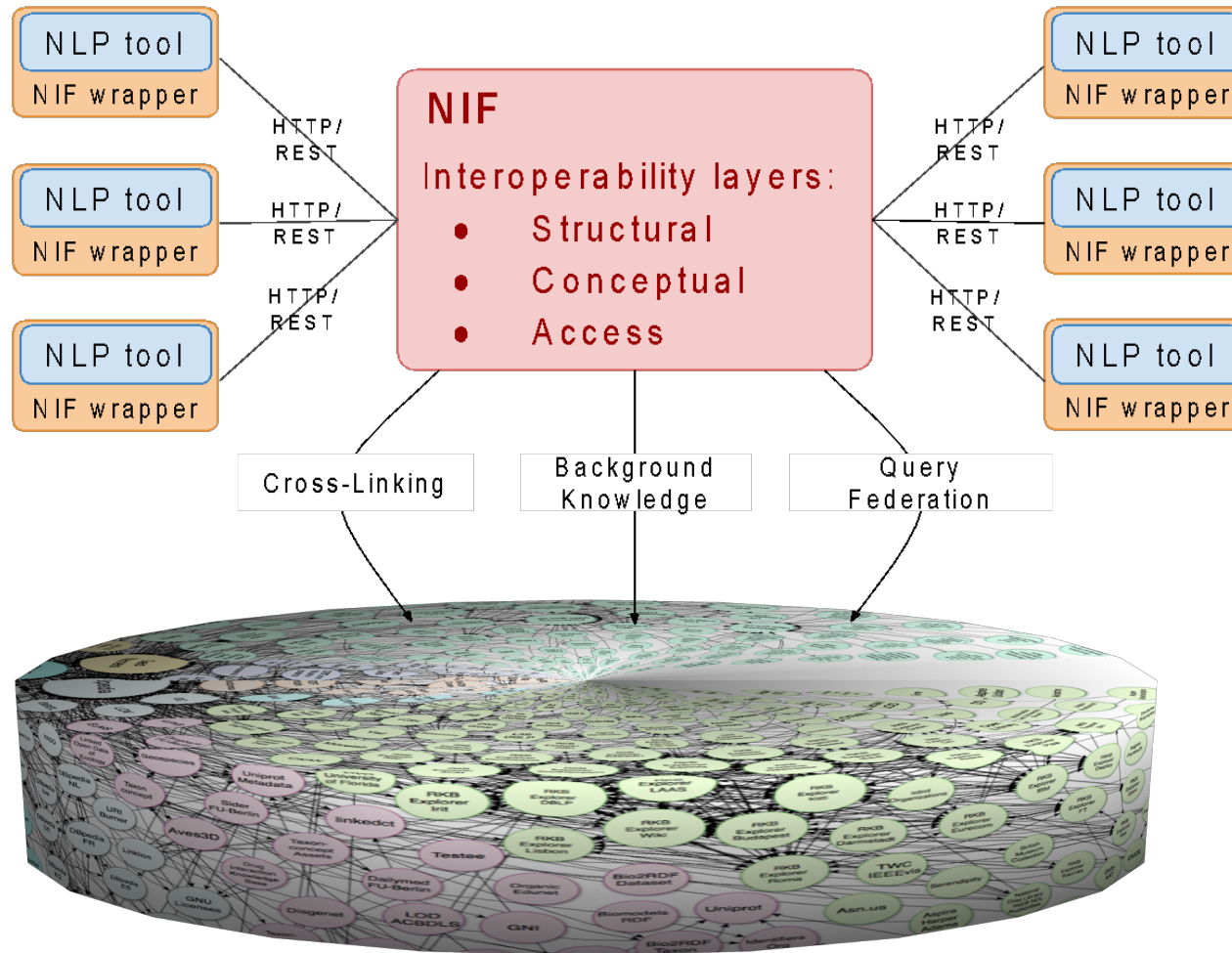
- One-to-one integration
- Hard to maintain



WTF! Spaghetti ?!!



How it works



NLP tools use one RDF vocabulary for information exchange

- NIF Wrapper

They refer to the same URIs (> strings)

Output of different tools can be merged

IF

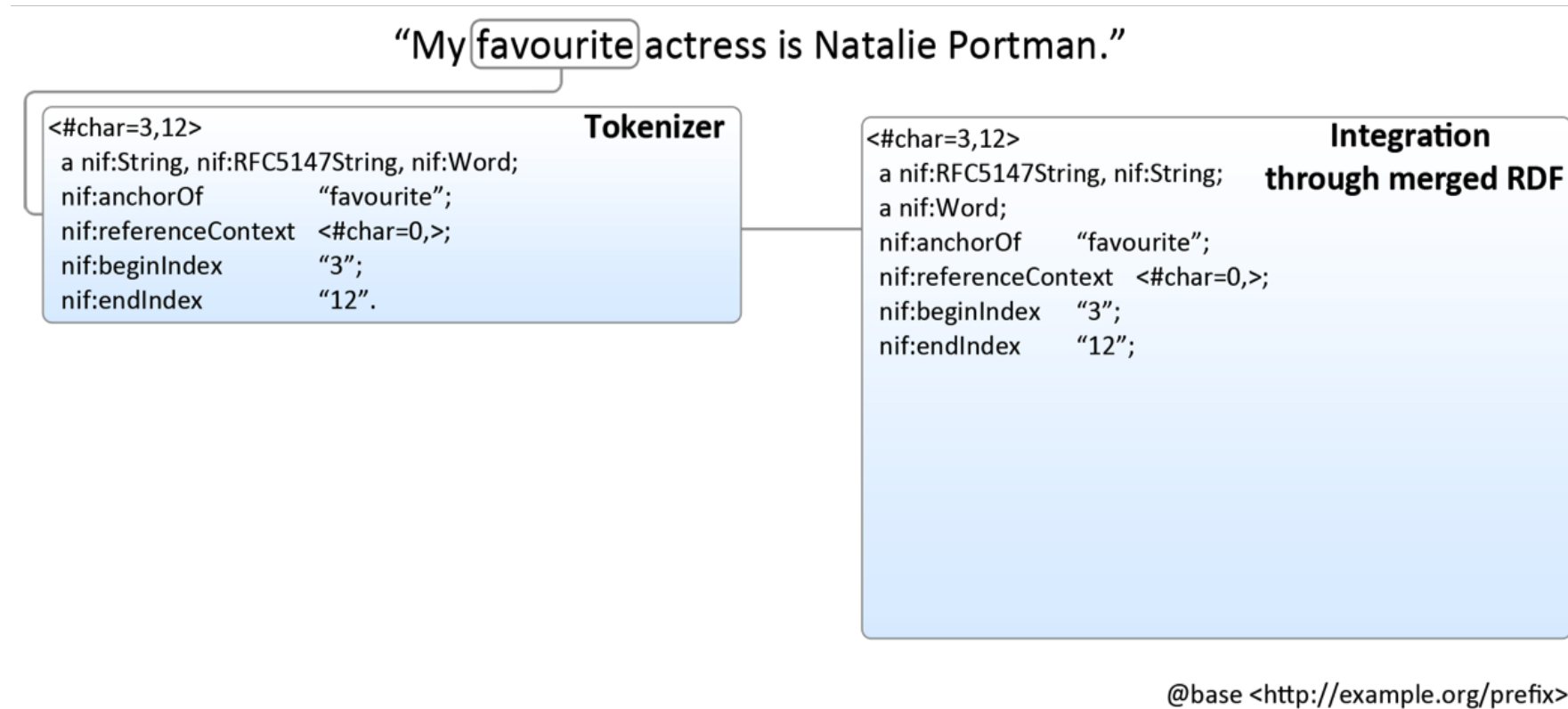
different tools produce the same URIs for the same string

AND

the annotation (RDF triples) is copied into the same RDF graph

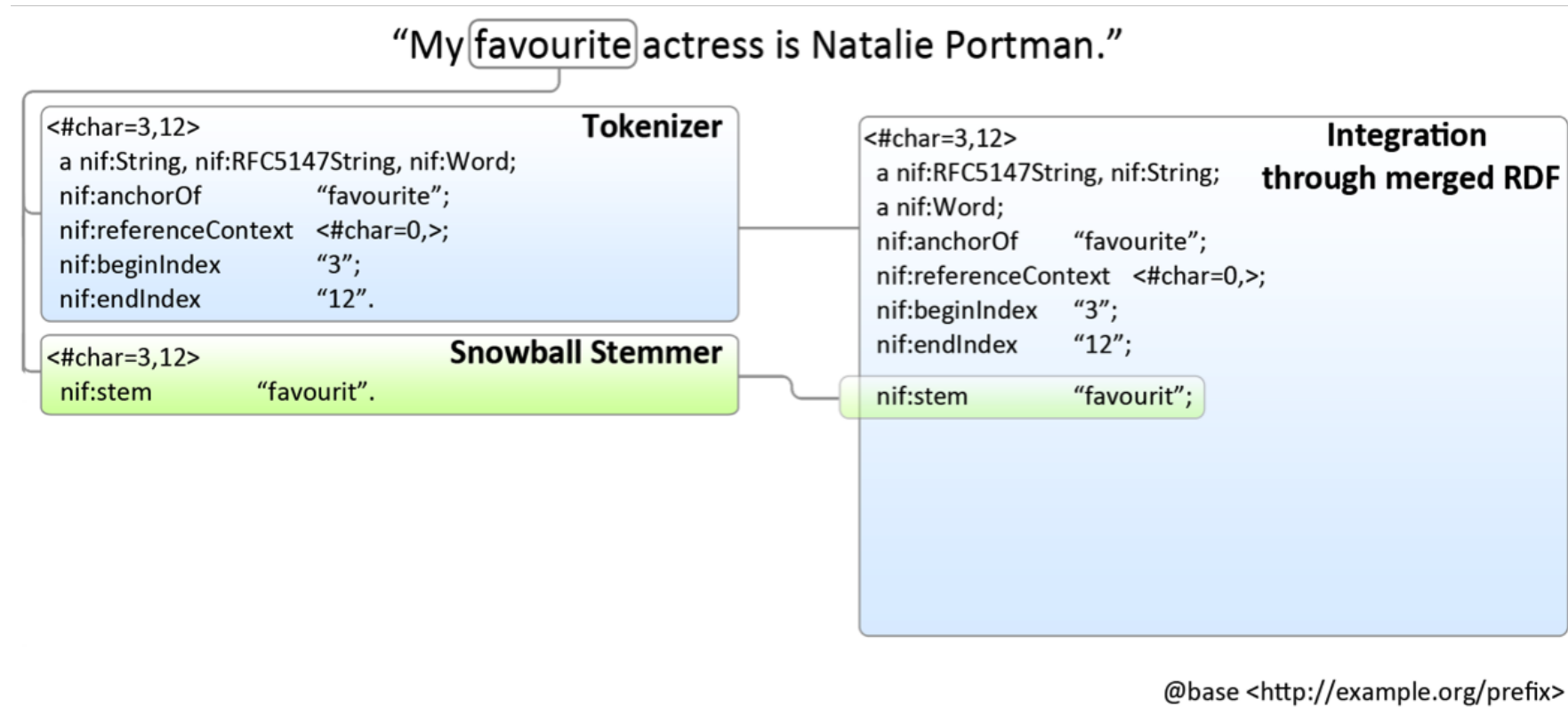


How it works: Simple tokenization



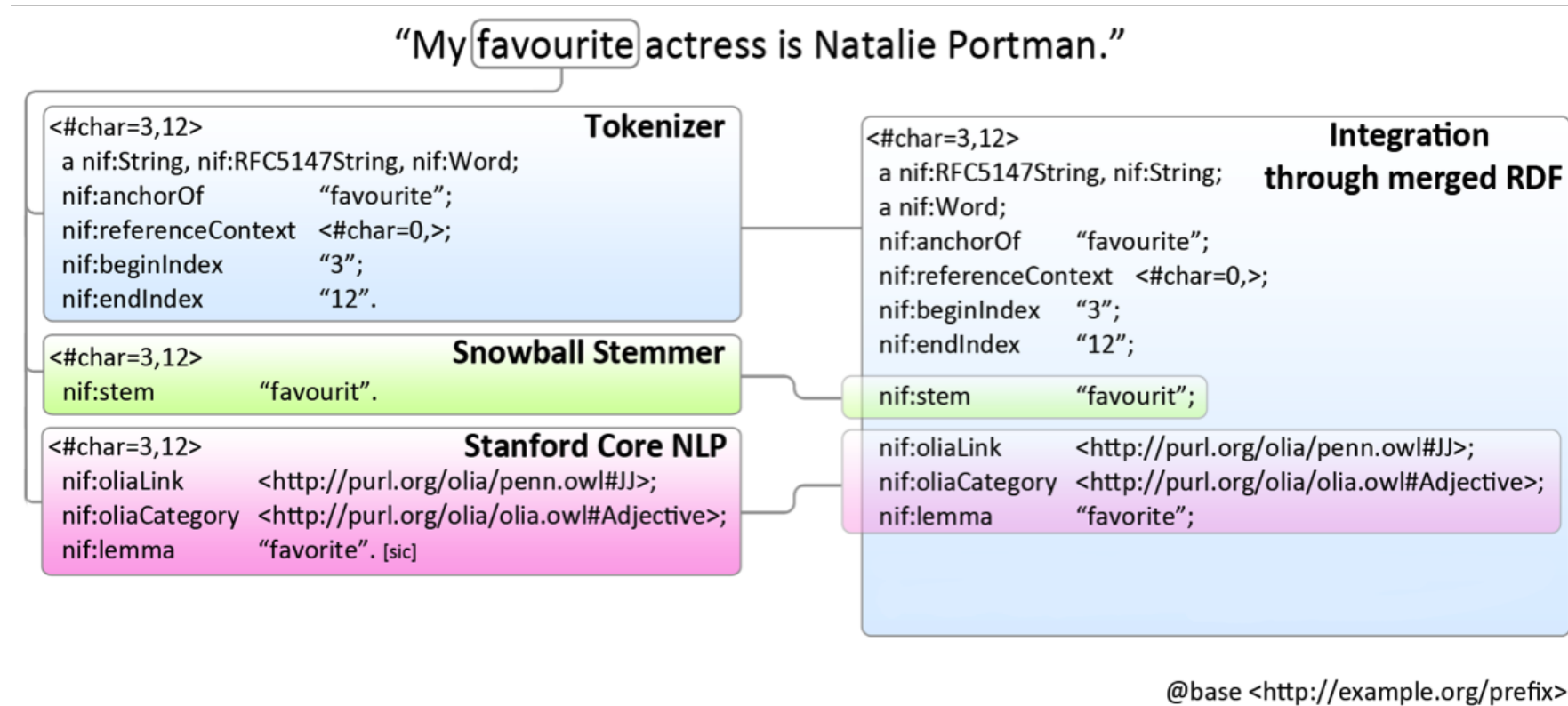


How it works: ... plus stemming



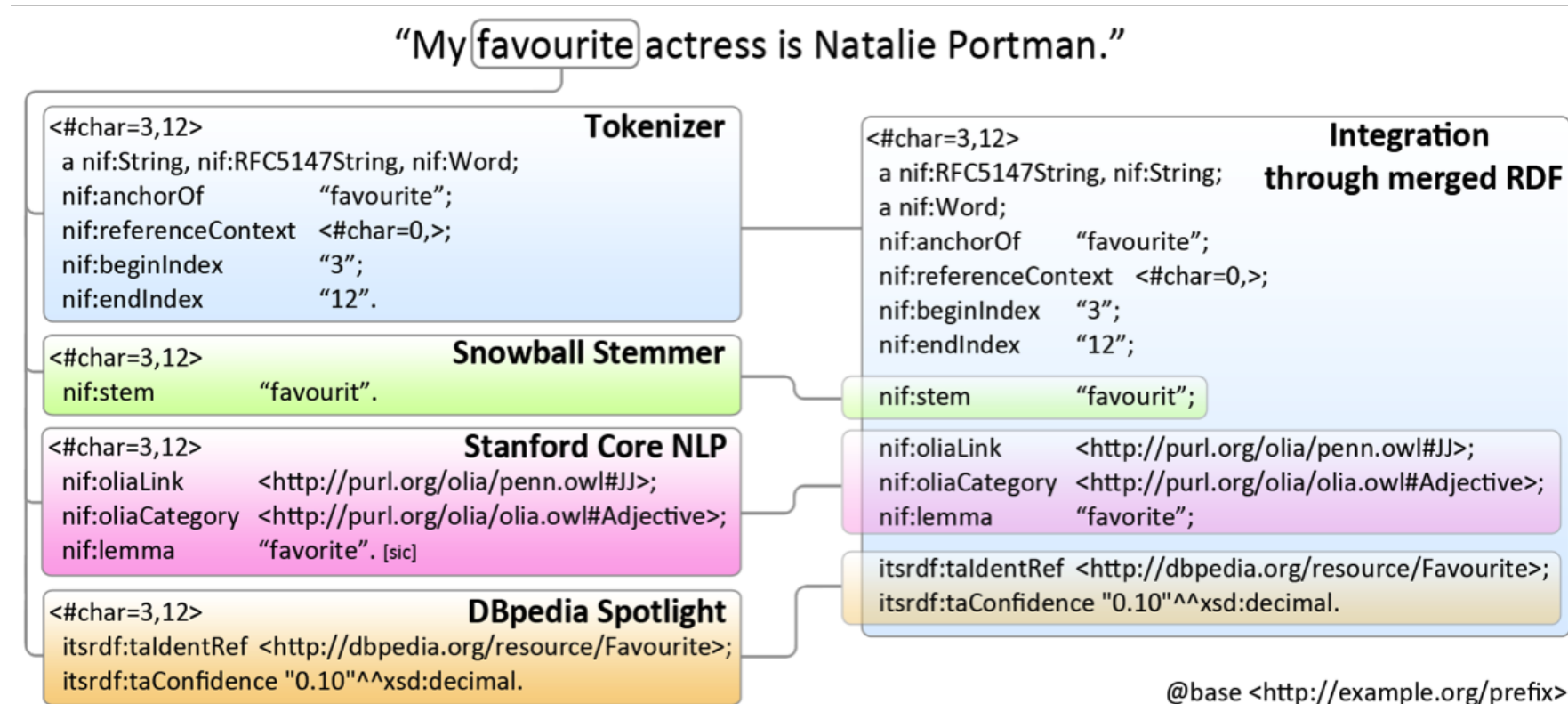


How it works: ... plus POS tags





How it works: ... plus entity linking

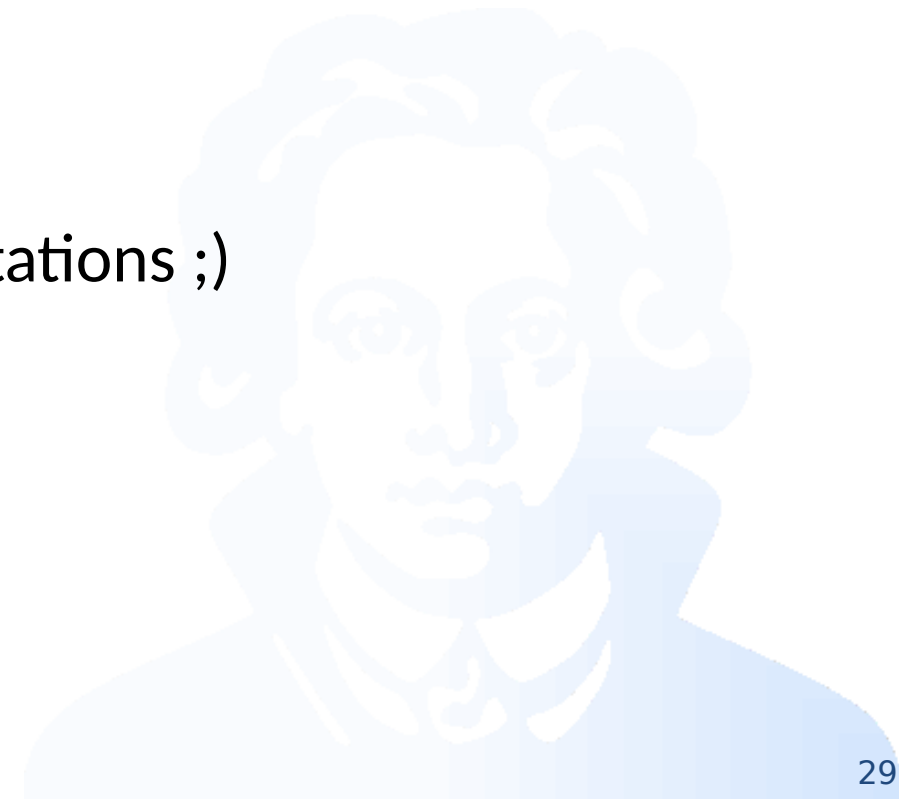




NIF - NLP Interchange Format

<http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core/>

- String URIs
 - e.g., in a web document
 - can be directly used as object of *oa:hasTarget*
- simple ontology of linguistic data structures
 - for selected, typical NLP annotations
 - *not* covering all you ever need for linguistic annotations ;)
- also contains specifications for NLP web services
 - advanced topic, not covered here





NIF String URIs

- RFC 5147 (for plain text documents)

- URI fragments for text references, via

- character position

- https://catalog ldc.upenn.edu/docs/LDC95T7/raw/06/ws_j_0655.txt#char=19

- character range

- https://.../ws_j_0655.txt#char=19,30





NIF String URIs

- RFC 5147 (for plain text documents)

- URI fragments for text references, via

- character position

- https://catalog ldc.upenn.edu/docs/LDC95T7/raw/06/wsj_0655.txt#char=19

- character range

- https://.../wsj_0655.txt#char=19,30

- Relative to a NIF context (optional)

- Offsets refer to the value of *nif:isString*:

- `<https://catalog ldc.upenn.edu/docs/LDC95T7/raw/06/wsj_0655.txt#char=19>`

- `nif:referenceContext [a nif:Context; nif:isString “this is your context string”] .`



NIF String URIs

- RFC 5147 (for plain text documents)

- URI fragments for text references, via

- character position

- https://catalog ldc.upenn.edu/docs/LDC95T7/raw/06/wsj_0655.txt#char=19

- character range

- https://.../wsj_0655.txt#char=19,30

- line offsets

- https://.../wsj_0655.txt#line=0 (first line)

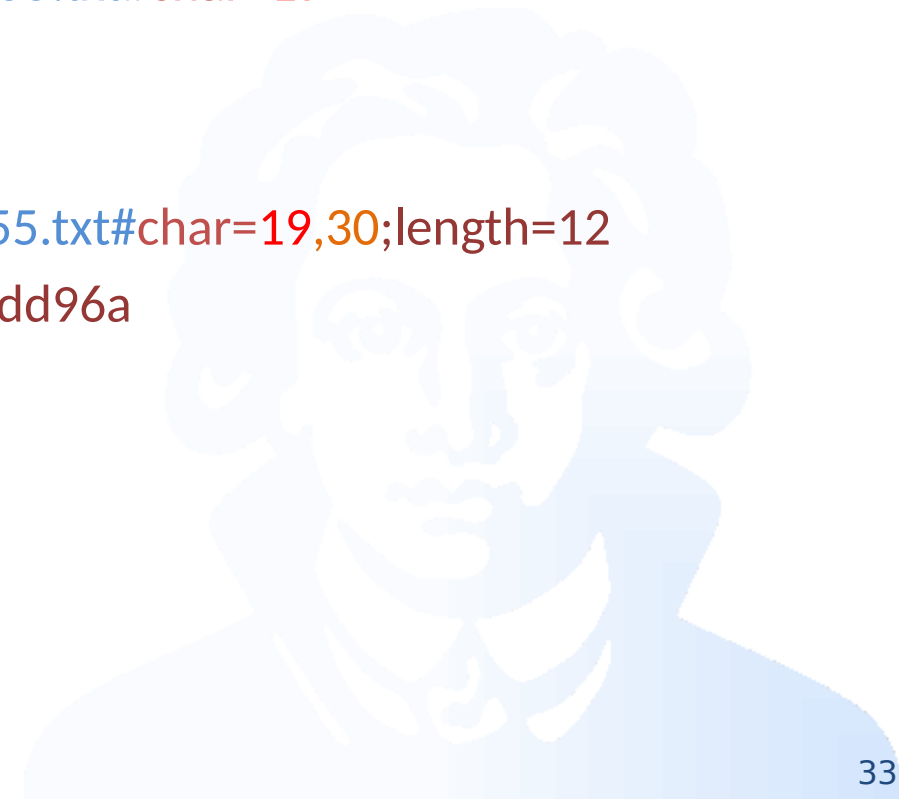
- https://.../wsj_0655.txt#line=0, (range until the end)





NIF String URIs

- RFC 5147 (for plain text documents)
 - URI fragments for text references, via
 - character position
 - https://catalog ldc.upenn.edu/docs/LDC95T7/raw/06/wsj_0655.txt#char=19
 - character range
 - https://.../wsj_0655.txt#char=19,30
 - optionally followed by integrity checks https://.../wsj_0655.txt#char=19,30;length=12
<https://...#char=19,30;md5=67f60186fe687bb898ab7faed17dd96a>





NIF String URIs

- RFC 5147 (for plain text documents)

- URI fragments for text references, via

- character position

- https://catalog ldc.upenn.edu/docs/LDC95T7/raw/06/wsj_0655.txt#char=19

- character range

- https://.../wsj_0655.txt#char=19,30

- optionally followed by integrity checks

- optionally followed by encoding specs

- https://.../wsj_0655.txt#char=19,30;length=12,UTF-8

- https://.../wsj_0655.txt#char=19,30;,UTF-8





NIF String URIs

- RFC 5147 (for plain text documents)

- URI fragments for text references, via

- character position

https://catalog ldc.upenn.edu/docs/LDC95T7/raw/06/ws_j_0655.txt#char=19

- character range

https://.../ws_j_0655.txt#char=19,30

```

8   oa:hasTarget [
9     oa:hasSelector [
10      a oa:TextQuoteSelector ;
11      oa:exact "James Baker"^^xsd:string
12    ] ;
13    oa:hasSource wsj:06/ws_j_0655.name
14  ] .

```

```

0   PREFIX wsj_0655 : <https://catalog ldc.upenn.edu/docs/LDC95T7/raw/06/ws_j_0655.txt#>
...

```

```

8-12   oa:hasTarget wsj_0655:char=19,30 .

```

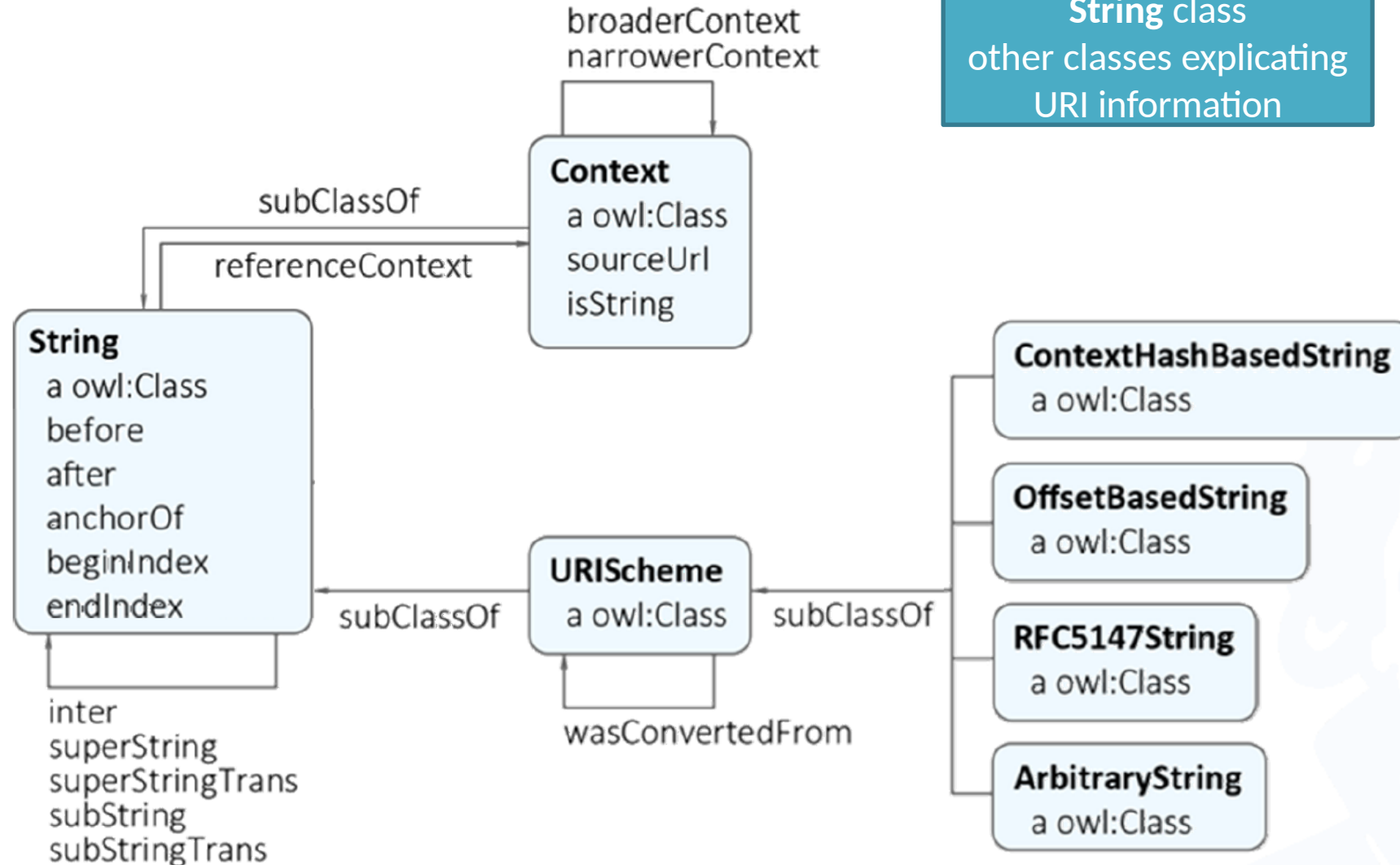
```

13-14  wsj_0655:char=19,30 oa:hasSource wsj:06/ws_j_0655.name .

```

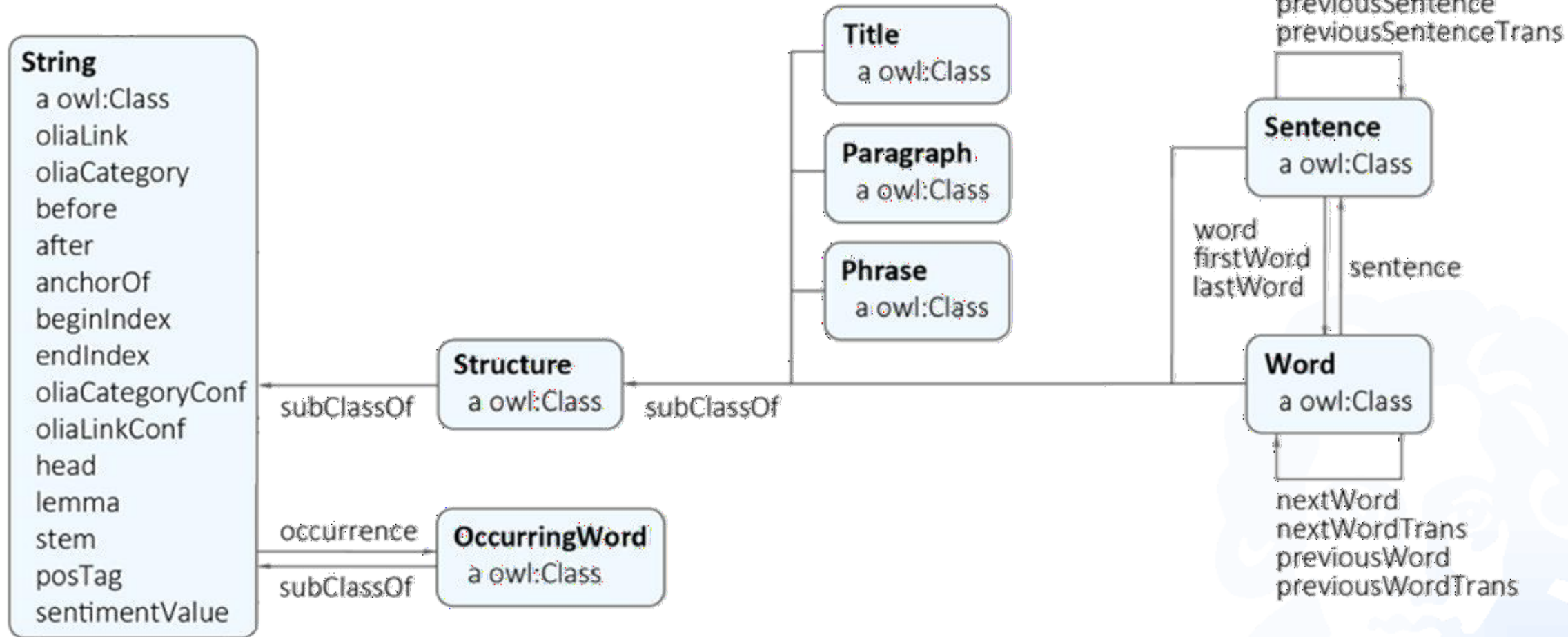


NIF 2.0 String (sub-)Ontology





NIF 2.0 Annotation (sub-)Ontology



subclasses of **String**
annotation properties for selected properties



Named Entity annotation in NIF (Turtle)

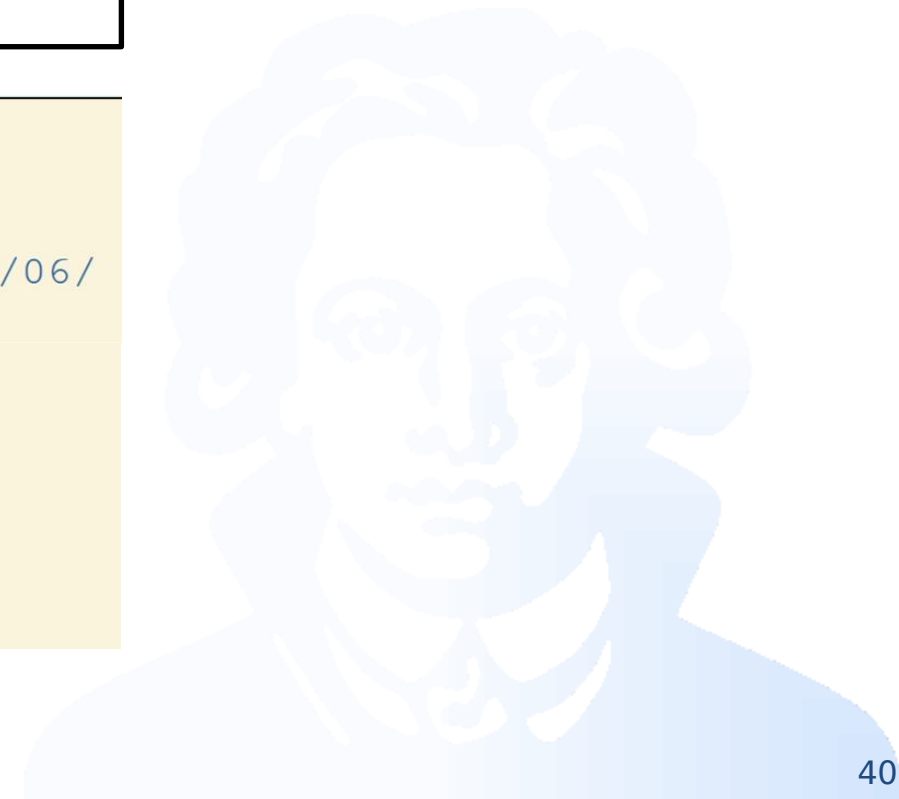
Secretary of State **James Baker**, who accompanied President **Bush** to **Costa Rica**, told reporters **Friday**: “I have no reason to deny reports that some **Contras** ambushed some **Sandinista** soldiers. ”

James	B-PERSON
Baker	E-PERSON
told	O
reporters	O
Friday	S-DATE
:	O

```

1 PREFIX nif: <http://persistence.uni-leipzig.org/nlp2rdf/
  ontologies/nif-core#>
2 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
3 PREFIX doc: <https://catalog.ldc.upenn.edu/docs/LDC95T7/raw/06/
  wsj_0655.txt#>
4 PREFIX on: <https://catalog.ldc.upenn.edu/docs/LDC2007T21/
  ontonotes-1.0-documentation.pdf#>
...
11 doc:offset_19_30
12     nif:anchorOf "James Baker";
13     nif:beginIndex "19" ; nif:endIndex "30" ;
14     on:ENAMEX "PERSON" .

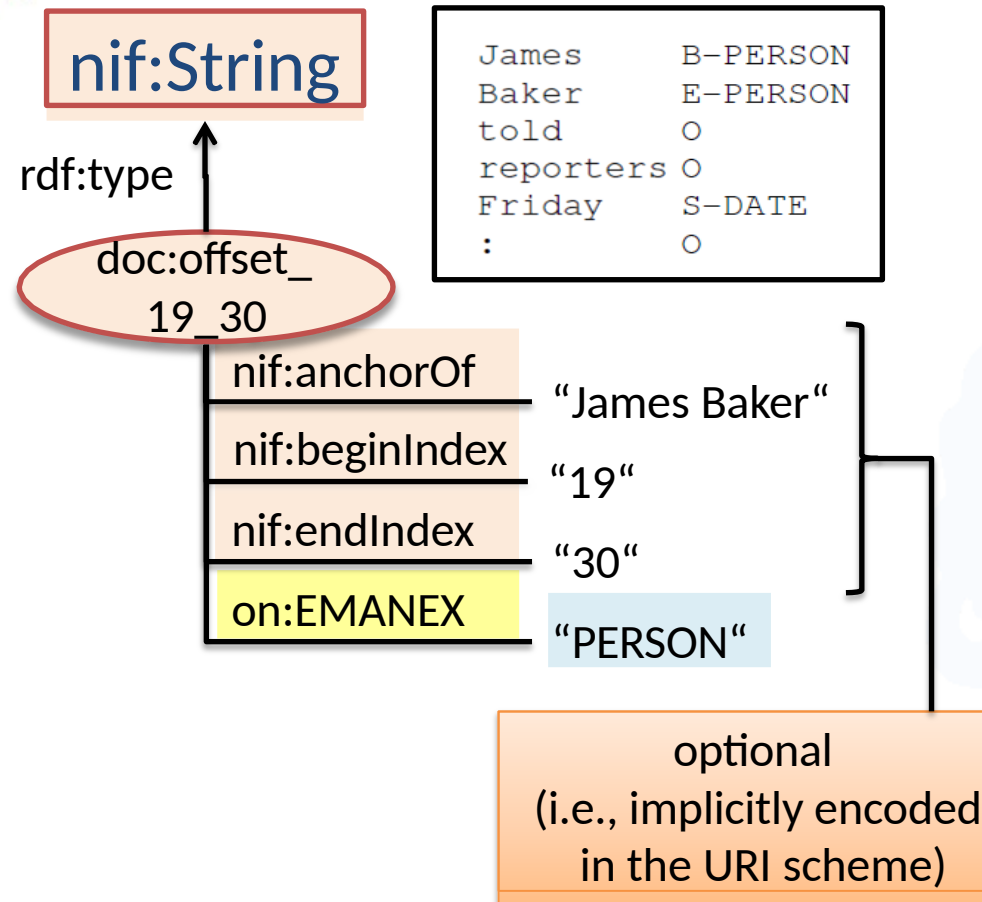
```





Named Entity annotation in NIF

Secretary of State James Baker, who accompanied President Bush to Costa Rica, told reporters Friday: “I have no reason to deny reports that some Contras ambushed some Sandinista soldiers. ”



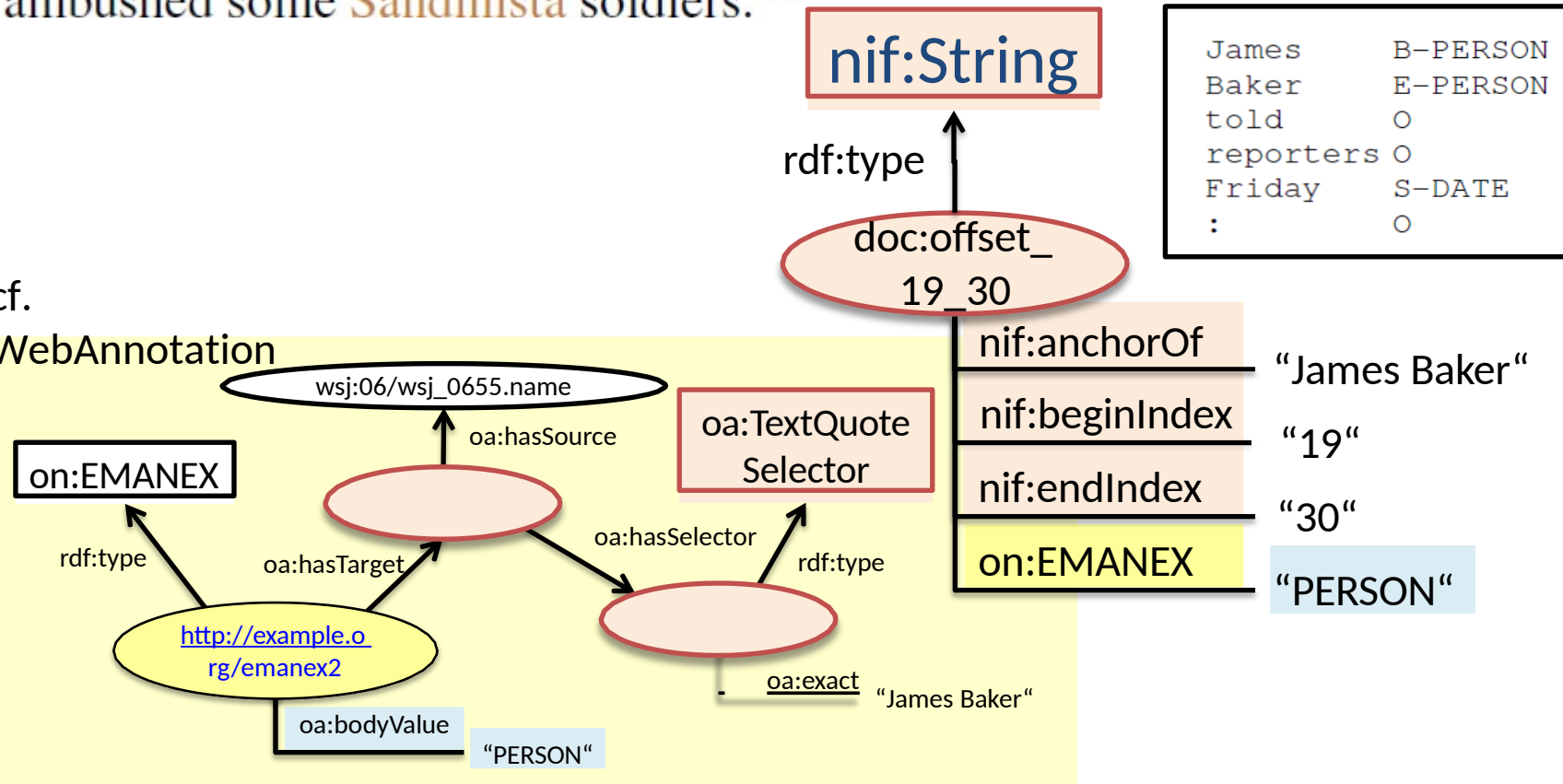


Named Entity annotation in NIF vs. WA

Secretary of State James Baker, who accompanied President Bush to Costa Rica, told reporters Friday: "I have no reason to deny reports that some Contras ambushed some Sandinista soldiers."

cf.

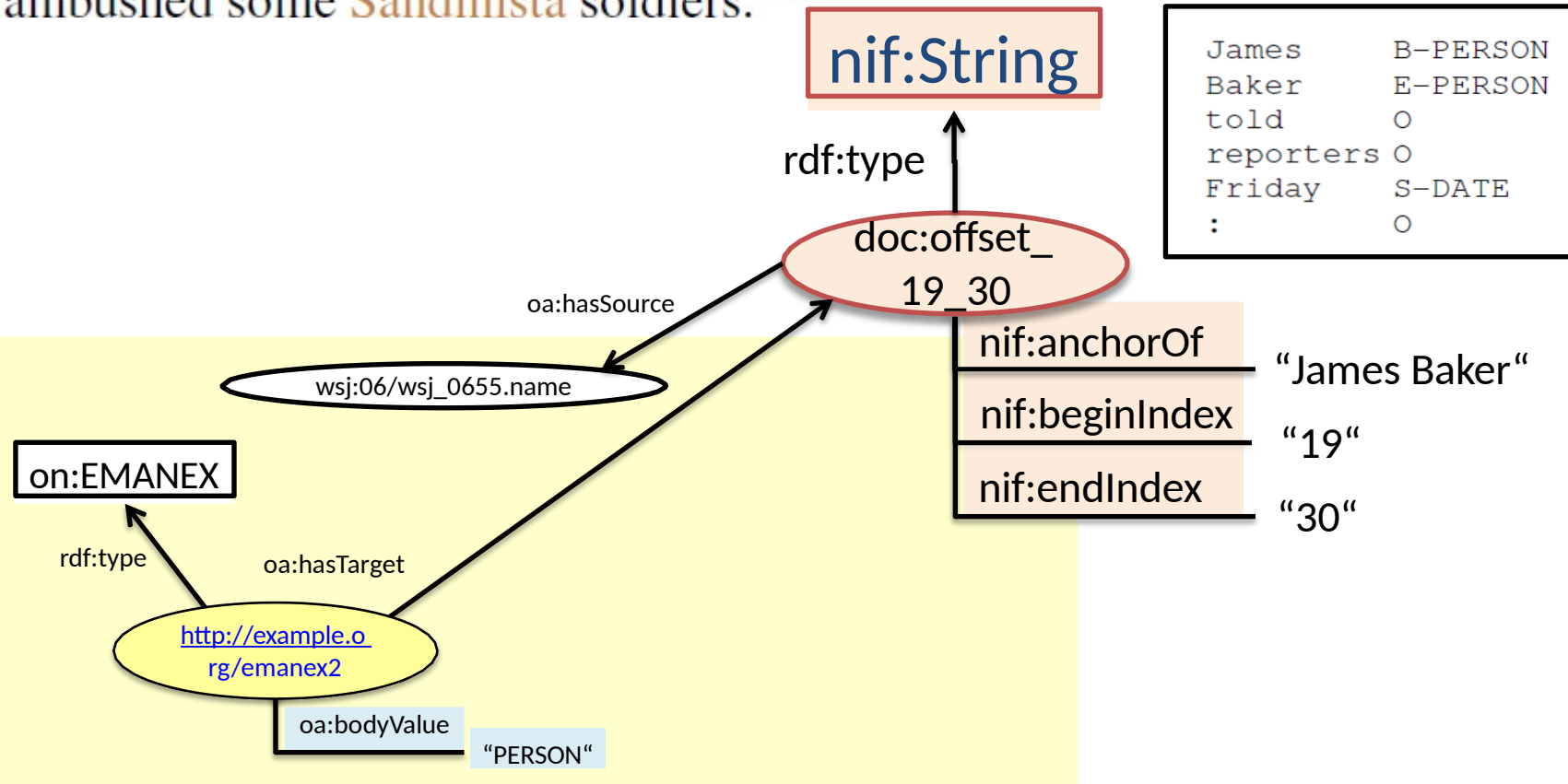
WebAnnotation





Named Entity annotation in NIF+WA

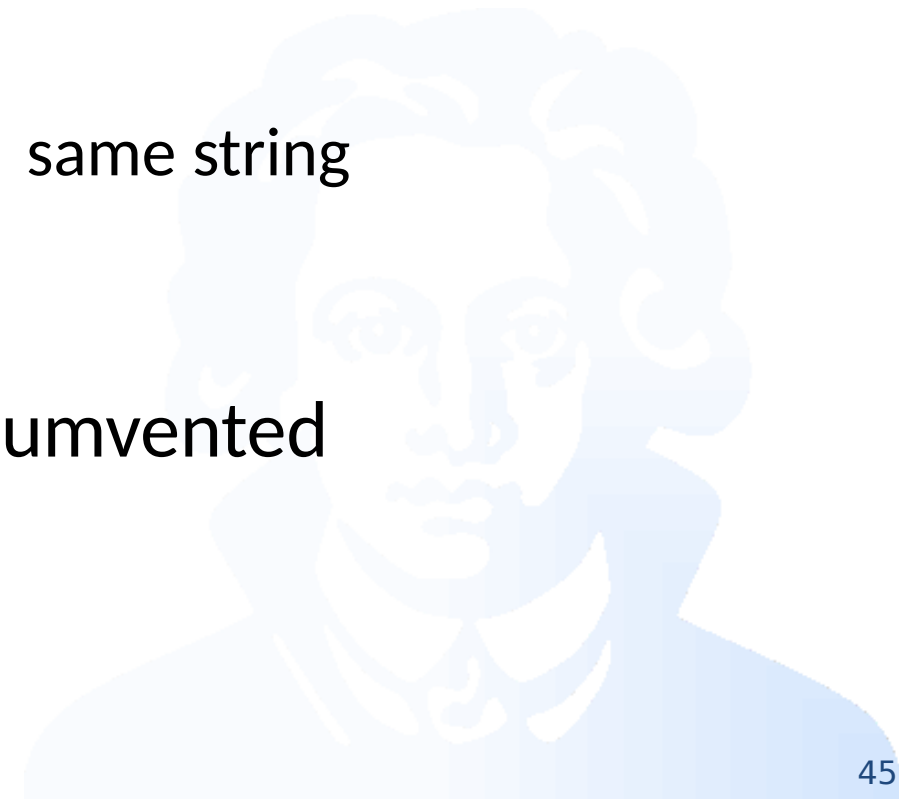
Secretary of State James Baker, who accompanied President Bush to Costa Rica, told reporters Friday: "I have no reason to deny reports that some Contras ambushed some Sandinista soldiers."





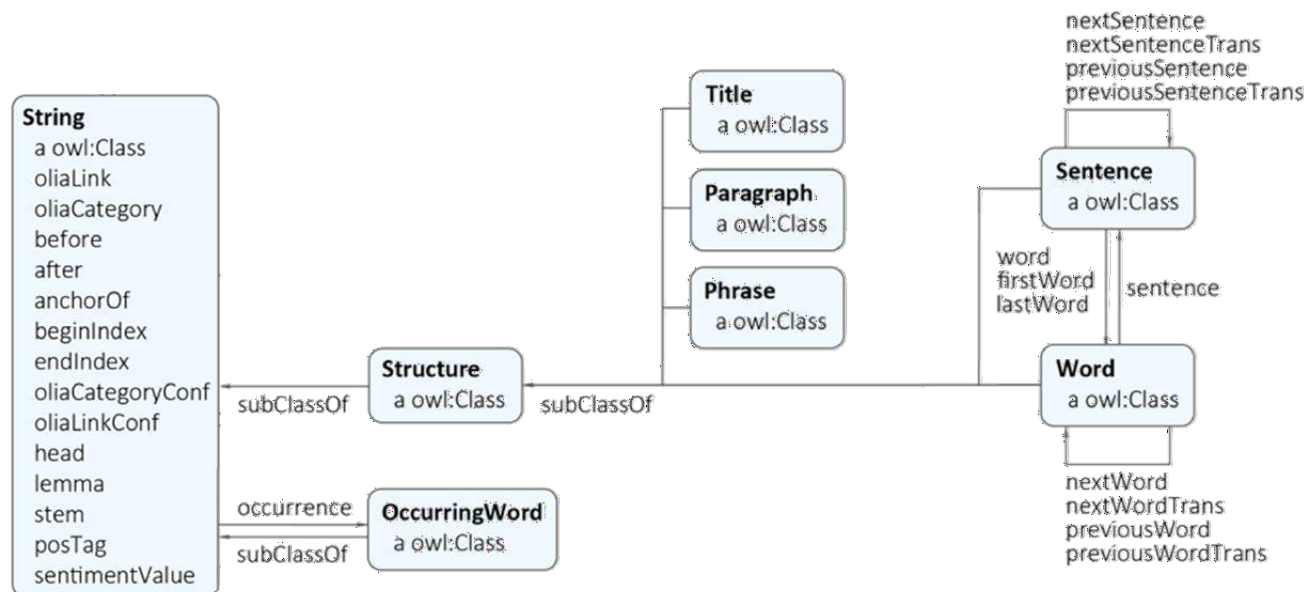
Using NIF

- widely known RDF-based annotation formalism for NLP
 - used, e.g., in Databus (<https://databus.dbpedia.org/>)
 - platform to provide global and unified access to data
- shallow, word-level annotations
 - implicit unification of annotations referring to the same string
 - useful in NLP
 - often unacceptable to linguists
- implicit unification of annotations can be circumvented
 - e.g., using Web Annotation



NIF limitations

- NIF data structures are no sufficient solution shallow for advanced linguistic annotations
 - ❑ morphology? (no sub-word strings)
 - ❑ non-phrasal/discontinuous segments ?
 - ❑ hard-wired annotation properties





 **Nexus**
Lingarum

4th Summer Datathon on Linguistic Linked Open Data
Cercedilla, Spain
30th May – 3rd June 2022

CoNLL-RDF

Making NIF (more) user-friendly





NIF Uptake

- NIF has been used for various webservices
 - But not in the NLP community nor language sciences
 - “What is that gibberish?”
 - “Never seen that in my training data!”
 - “Are you kidding? My data is TSV, my tools write TSV!”
 - “How can I feed *that* into *anything*?”
 - “How to I get there from my corpora?”
 - “I have X in my data. Is that even feasible with NIF?”
- => CoNLL-RDF (Chiarcos & Fäth 2017, Chiarcos et al. 2021)



CoNLL format family

- CoNLL is a format family widely used in NLP
 - tab-separated values
 - one word per line
 - one column for annotation type
 - sentences separated by empty lines
 - conventions for most types of word-based linguistic annotation
 - SRL, dependency syntax, phrase structure syntax, span annotation, coreference
- and beyond
 - cf. SketchEngine, NoSketchEngine, CorpusWorkbench
 - one-word-per-line TSV formats

James	B-PERSON
Baker	E-PERSON
told	O
reporters	O
Friday	S-DATE
:	O



CoNLL-RDF: idea

- isomorphic rendering of CoNLL data structures in RDF
 - every line a *nif:Word*
 - every block a *nif:Sentence*
- user provides labels for column names
 - column label => property in the *conll:* namespace
 - no underlying ontology
 - columns *HEAD* and *SRL-ARGs* become object properties
 - „foreign keys“, references to other words
 - other columns become datatype properties
 - annotation as literal value



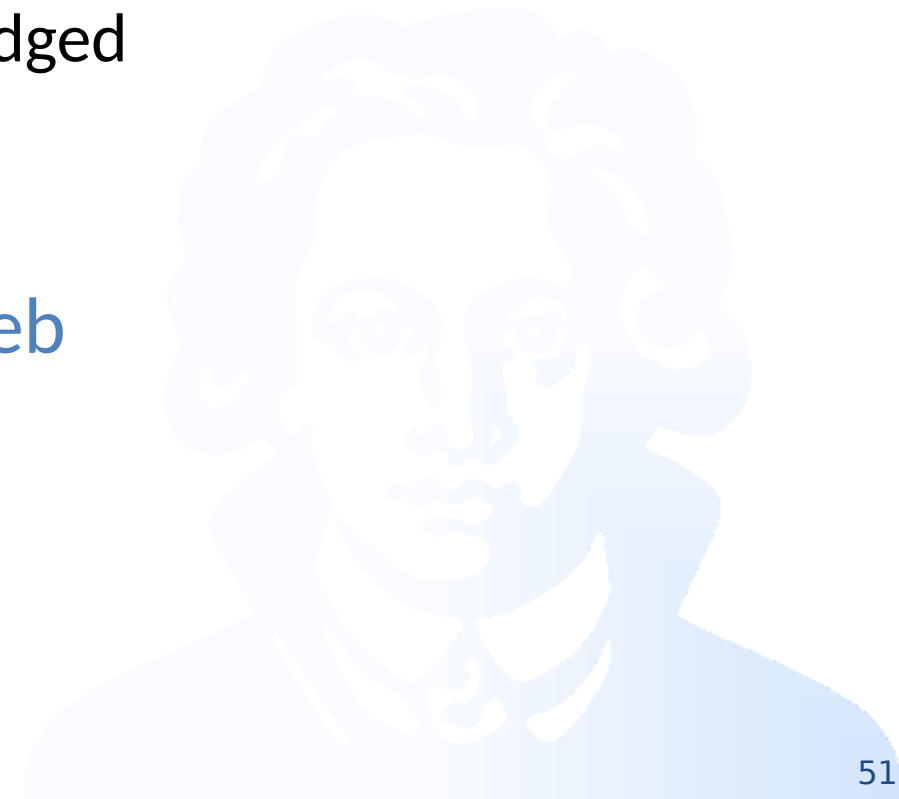


CoNLL-RDF: motivation

■ Why?

- ❑ create an RDF graph for every sentence
- ❑ use SPARQL Update for manipulating or enriching annotations
- ❑ serialize back into CoNLL, other formats or full-fledged Linked Data

– seamless integration with existing TSV-based workflows/corpus tech **and** with Semantic Web technologies

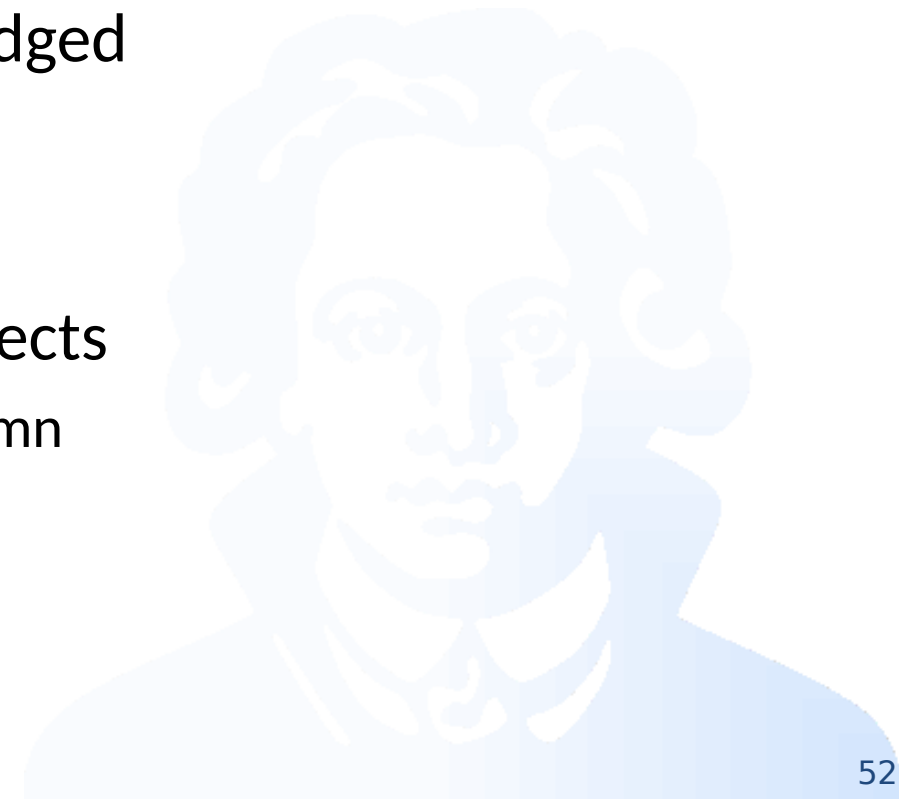




CoNLL-RDF: motivation

■ Why?

- ❑ create an RDF graph for every sentence
 - ❑ use SPARQL Update for manipulating or enriching annotations
 - ❑ serialize back into CoNLL, other formats or full-fledged Linked Data
- ## ■ decouples format from transformation logic
- ❑ transformations applicable to different CoNLL dialects
 - we don't care whether column *EDGE* is 3rd or 42th column





CoNLL-RDF components <https://github.com/acoli-repo/conll-rdf>

- CoNLL-RDF library (Chiacos & Fäth 2017)
 - Convert CoNLL to RDF (CoNLL-RDF)
 - Split into sentences and apply SPARQL transformations
 - Serialize back into CoNLL





CoNLL-RDF components <https://github.com/acoli-repo/conll-rdf>

- CoNLL-RDF library (Chiacos & Fäth 2017)
 - Convert CoNLL to RDF (CoNLL-RDF)
 - Split into sentences and apply SPARQL transformations
 - Serialize back into CoNLL

- CoNLL-RDF vocabulary (Chiacos et al. 2021)
 - Mirror CoNLL data structures in RDF
 - one *nif:Word* per row
 - one *conll:XYZ* property per column (user-provided column label XYZ)
 - one *nif:Sentence* per \n\n-separated block



CoNLL-RDF components <https://github.com/acoli-repo/conll-rdf>

- CoNLL-RDF library (Chiarcos & Fäth 2017)
 - Convert CoNLL to RDF (CoNLL-RDF)
 - Split into sentences and apply SPARQL transformations
 - Serialize back into CoNLL

- CoNLL-RDF vocabulary (Chiarcos et al. 2021)
 - Special extensions for
 - syntactic dependencies (*conll:HEAD*)
 - semantic role labelling (*conll:ARG1*)
 - phrase structure annotations (*powla:Node*, *powla:Relation*)



CoNLL-RDF components <https://github.com/acoli-repo/conll-rdf>

- CoNLL-RDF library (Chiarcos & Fäth 2017)
 - ➔ FINTAN (Fäth et al. 2020)
 - Consume various input formats (incl. CoNLL)
 - Transform with SPARQL and serialize (incl. CoNLL)
 - Workflow manager
- CoNLL-RDF vocabulary (Chiarcos et al. 2021)
 - Special extensions for
 - syntactic dependencies (*conll:HEAD*)
 - semantic role labelling (*conll:ARG1*)
 - phrase structure annotations (*powla:Node*, *powla:Relation*)



CoNLL-RDF example

- Input: TSV format
- Parameter: Column labels
 - Here: WORD NER
- CoNLL-RDF (FINTAN) CLI

James	B-PERSON
Baker	E-PERSON
told	O
reporters	O
Friday	S-DATE
:	O

```
cat sample.conll | ~/conll-rdf/run.sh CoNLLStreamExtractor '#' WORD NER
```

```
:s1_0 a nif:Sentence .
:s1_1 a nif:Word; conll:WORD "James"; conll:HEAD :s1_0; conll:NER "B-Person"; nif:nextWord :s1_2 .
:s1_2 a nif:Word; conll:WORD "Baker"; conll:HEAD :s1_0; conll:NER "E-Person"; nif:nextWord :s1_3 .
:s1_3 a nif:Word; conll:WORD "told"; conll:HEAD :s1_0; nif:nextWord :s1_4 .
:s1_4 a nif:Word; conll:WORD "reporters"; conll:HEAD :s1_0; nif:nextWord :s1_5 .
:s1_5 a nif:Word; conll:WORD "Friday"; conll:HEAD :s1_0; conll:NER "S-DATE"; nif:nextWord :s1_6 .
:s1_6 a nif:Word; conll:WORD ":"; conll:HEAD :s1_0 .
```



CoNLL-RDF example

- Input: TSV format
- Parameter: Column labels
 - Here: WORD NER
- CoNLL-RDF (FINTAN) CLI

James	B-PERSON
Baker	E-PERSON
told	O
reporters	O
Friday	S-DATE
:	O

```
cat sample.conll | ~/conll-rdf/run.sh CoNLLStreamExtractor '#' WORD NER | \
~/conll-rdf/run.sh CoNLLRDFFormatter -conll WORD NER
```

```
# global.columns = WORD NER
James      B-Person
Baker      E-Person
told       -
reporters  -
Friday     S-DATE
:          -
```



CoNLL-RDF example

- Input: TSV format
- Parameter: Column labels
 - Here: WORD NER
- CoNLL-RDF (FINTAN) CLI

James	B-PERSON
Baker	E-PERSON
told	O
reporters	O
Friday	S-DATE
:	O

```
cat sample.conll | ~/conll-rdf/run.sh CoNLLStreamExtractor '#' WORD NER | \  
~/conll-rdf/run.sh CoNLLRDFUpdater -custom YOUR.sparql | \  
~/conll-rdf/run.sh CoNLLRDFFormatter -conll WORD NEW_COL
```

YOUR.sparql: whatever SPARQL transformation you want to apply to your data
NEW_COL: whatever (new) property in the conll: namespace you want to export to CoNLL

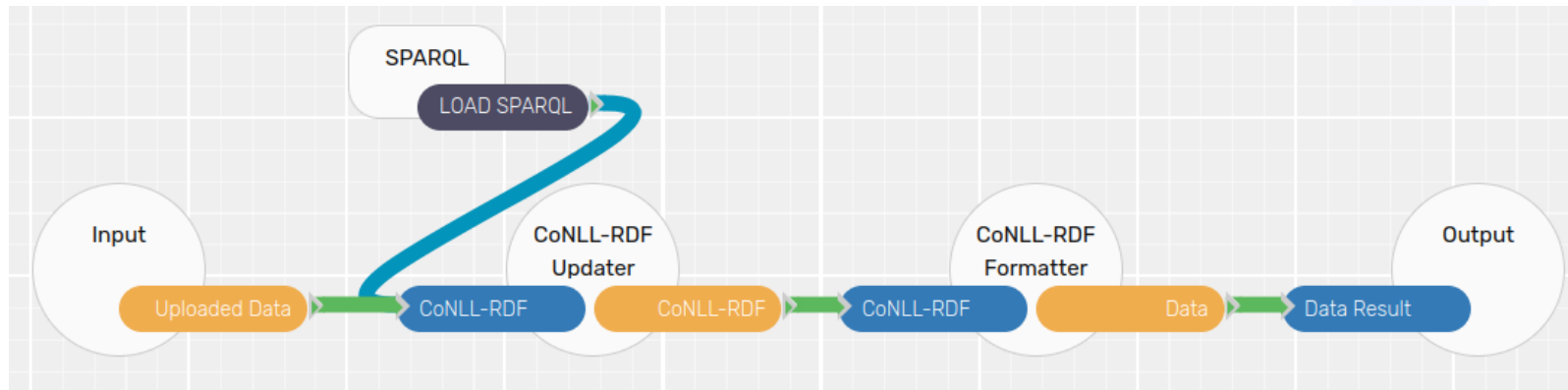


CoNLL-RDF example

- Input: TSV format
- Parameter: Column labels
 - Here: WORD NER
- (CoNLL-RDF) FINTAN

James	B-PERSON
Baker	E-PERSON
told	O
reporters	O
Friday	S-DATE
:	O

<https://github.com/Pret-a-LLOD/Fintan> for installation instructions



YOUR.sparql: whatever SPARQL transformation you want to apply to your data
NEW_COL: whatever (new) property in the conll: namespace you want to export to CoNLL



 **Nexus**
Lingarum

4th Summer Datathon on Linguistic Linked Open Data
Cercedilla, Spain
30th May – 3rd June 2022

Advanced Topics



Advanced: Complementary Vocabularies

■ POWLA (Chiarcos 2012)

<http://purl.org/powla/powla.owl>

- Data structures for
 - phrases (*powla:Node*)
 - relations (*powla:Relation*)

■ OLiA (Ontologies of Linguistic Annotation, Chiarcos & Sukhareva 2015)

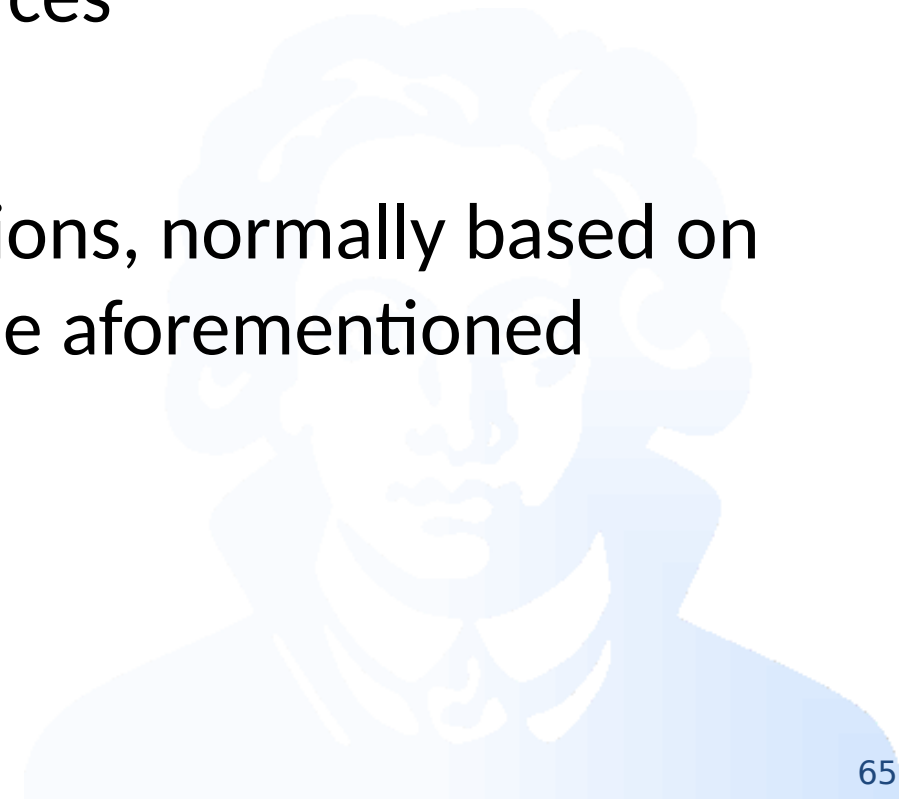
<http://purl.org/olia>

- Machine-readable annotation schemes (tagsets) in OWL
- Linked with reference terminology (OLiA Reference Model)



Advanced: Web Services

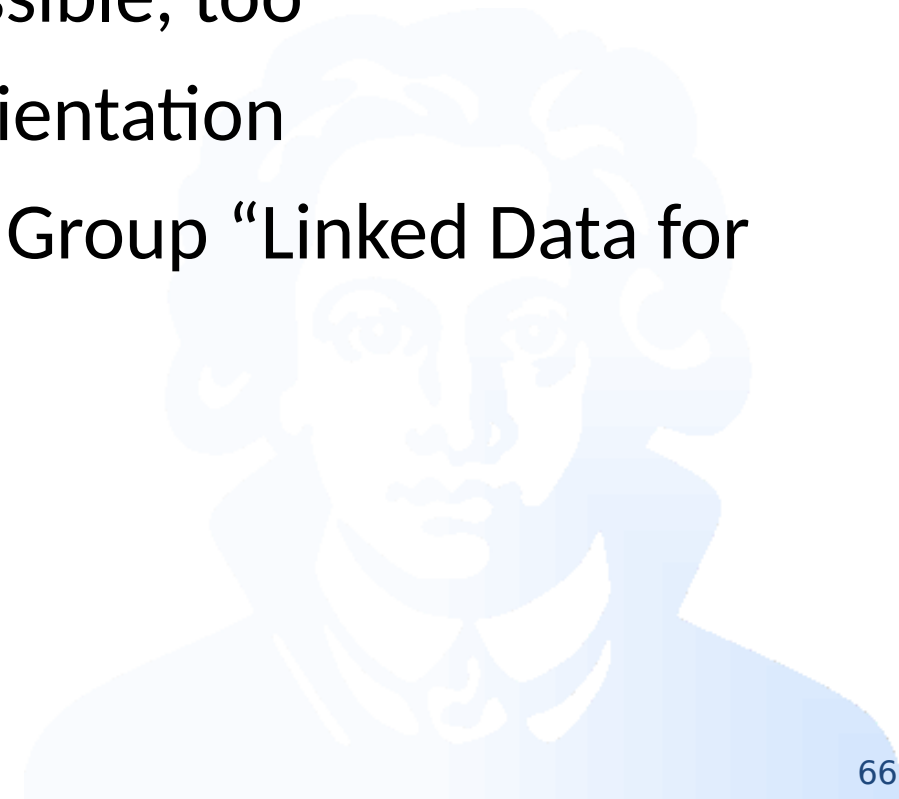
- *incompatible* specifications as part of NIF and Web Annotation
 - NIF: ideal for annotation words out of context, e.g., if query responses are annotated
 - Web Annotation: ideal for static web resources
- NLP workflows
 - often using system-specific/ad hoc specifications, normally based on REST and using a system-specific or one of the aforementioned formats
 - Teanga, OpenMinTeD, LAPPS Grid, etc.





Challenge: Harmonizing Community Standards

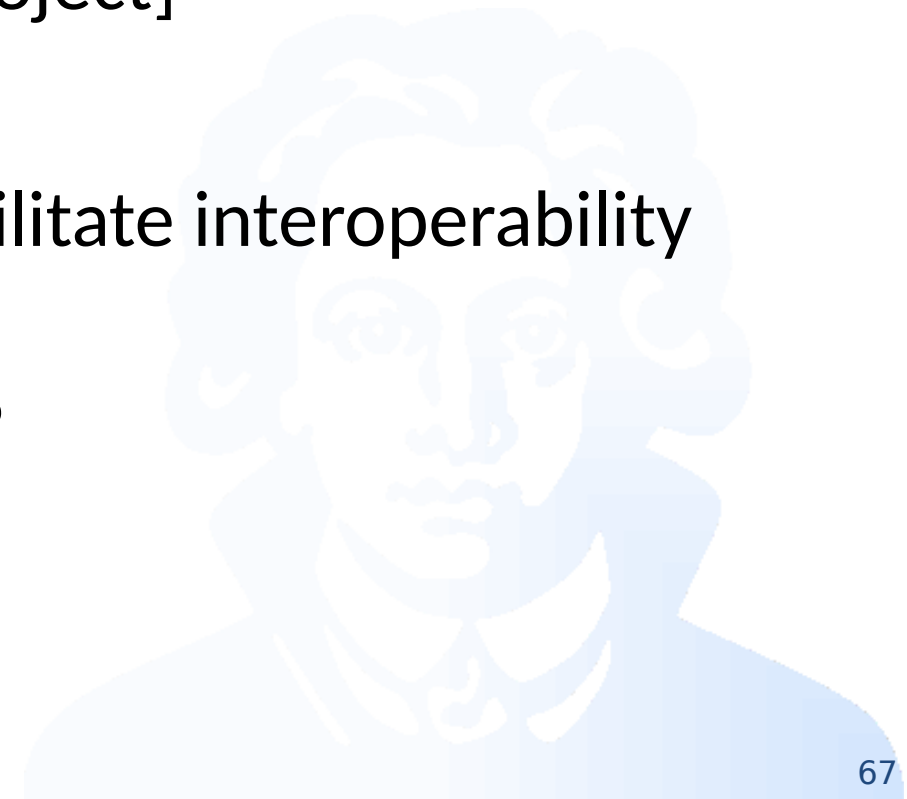
- For lexical resources, the lack of alternatives to OntoLex has led to enormous synergies
 - This is now mainstream technology, with wide adaptation
- For linguistic annotation, such an effect is possible, too
 - Curbed by incompatibilities and a lack of orientation
 - Consolidation initiative at W3C Community Group “Linked Data for Language Technology” (LD4LT)
 - <https://github.com/ld4lt/linguistic-annotation>
 - <https://www.w3.org/community/ld4lt/>





Challenges: Tool support

- LLOD-native tools are mostly concerned with
 - entity linking, knowledge extraction, lexicography
 - NEW: processing workflows [Pret-a-LLOD project]
- Few end user tools for *linguistic annotation*
 - at the moment, RDF is mostly applied to facilitate interoperability between legacy tools and data
 - backend-centered: “an invisible technology”





Acknowledgments

Partially based on presentations from SD-LLOD 2019, these slides were created for SD-LLOD 2022, Madrid, Jun 2022, by Christian Chiarcos, Applied Computational Linguistics, GU Frankfurt, Germany

Use with the following attribution

- Christian Chiarcos, *Annotations and Natural Language Processing*, seminar at SD-LLOD 2022, Madrid, June 2022
- Licensed as [CC BY-NC-SA 3.0 Unported](https://creativecommons.org/licenses/by-nc-sa/3.0/)