



Language Resources and Linked Data

Integrating NLP with Linked Data: the NIF Format

@EKAU 2014

November 24-28, 2014, Linköping, Sweden

Milan Dojchinovski

milan.dojchinovski@fit.cvut.cz - @m1ci - <http://dojchinovski.mk>



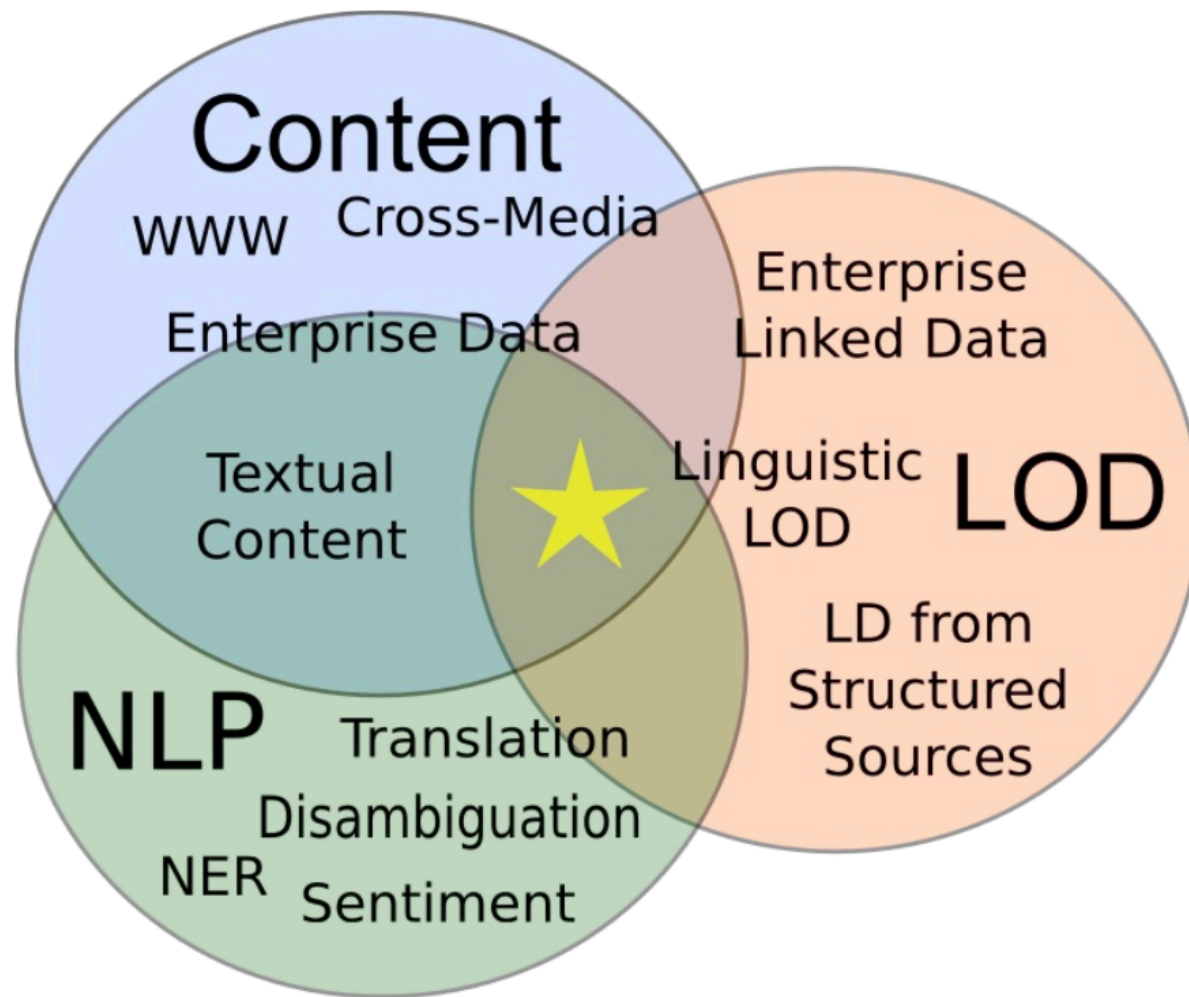
Web Intelligence Research Group
Faculty of Information Technology
Czech Technical University in Prague

1. Introduction

- NIF Basics
- NIF Corpora
- NIF Tools and Services

2. Hands-on: NIF in action

- How to annotate strings
- How to query generated NIF and existing corpora



- Not only data, but also LOD-aware services using:
 - Lexica and dictionaries (lemon model)
 - Training data for NLP in RDF (NIF model)
 - Service metadata descriptions in RDF
 - Combination with real world facts (i.e. DBpedia or GeoNames)
- Long term goal(s):
 - Index of tools and data
 - Easily produce ready-made, preconfigured NLP services and pipelines
 - freemium /pay-per-use business models

- Maintained under <http://nlp2rdf.org>
- Realize the long term goal(s)
- Maintain and consolidate results from short-term projects
- Bootstrap the eco-system

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

- Way to mint URIs for arbitrary strings and content documents on the Web
- Logical formalisation of strings and annotations via an ontology
- Quick and easy format
- Build on existing standards (RDF, LAF/GrAF, RFC 5147)
- Reuse of RDF tools and implementations
- Decrease development costs for integration

- 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 way of annotating text
 - HTML markup removed – offsets not usable
 - Use of byte offset instead of char offset

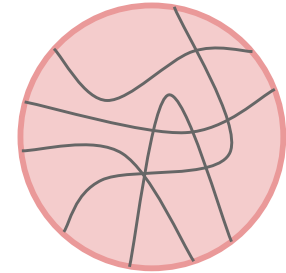
1. Introduction

- **NIF Basics**
- NIF Corpora
- NIF Tools and Services

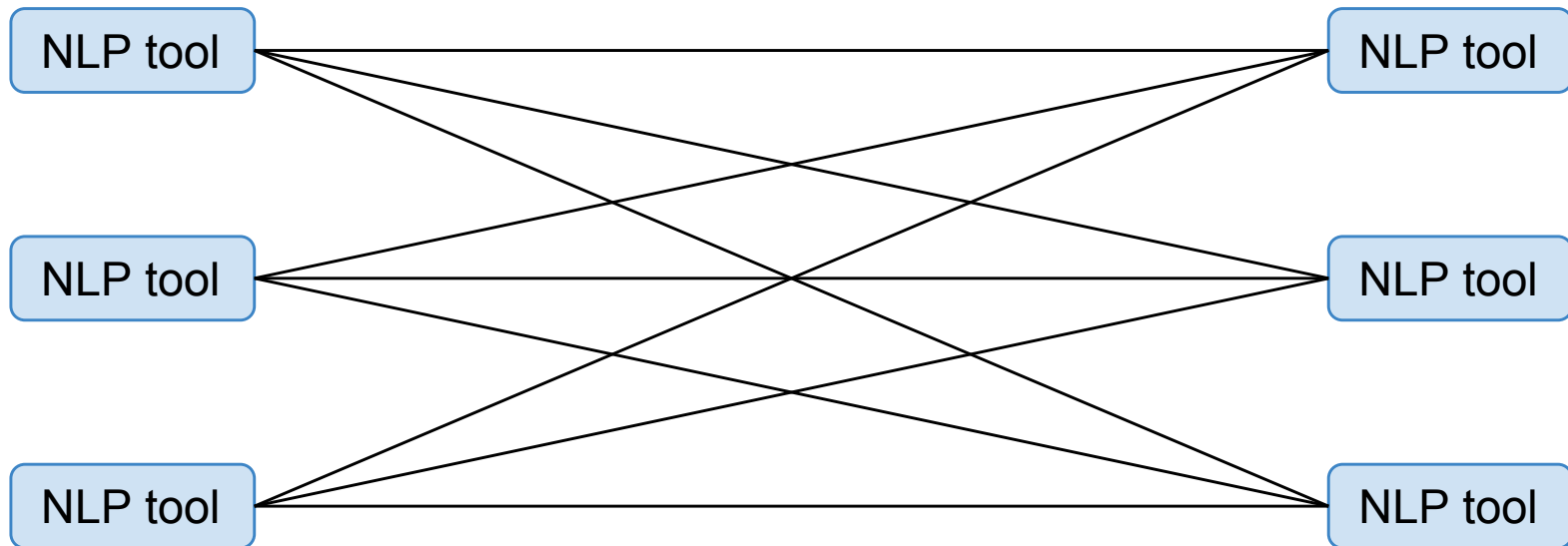
2. Hands-on: NIF in action

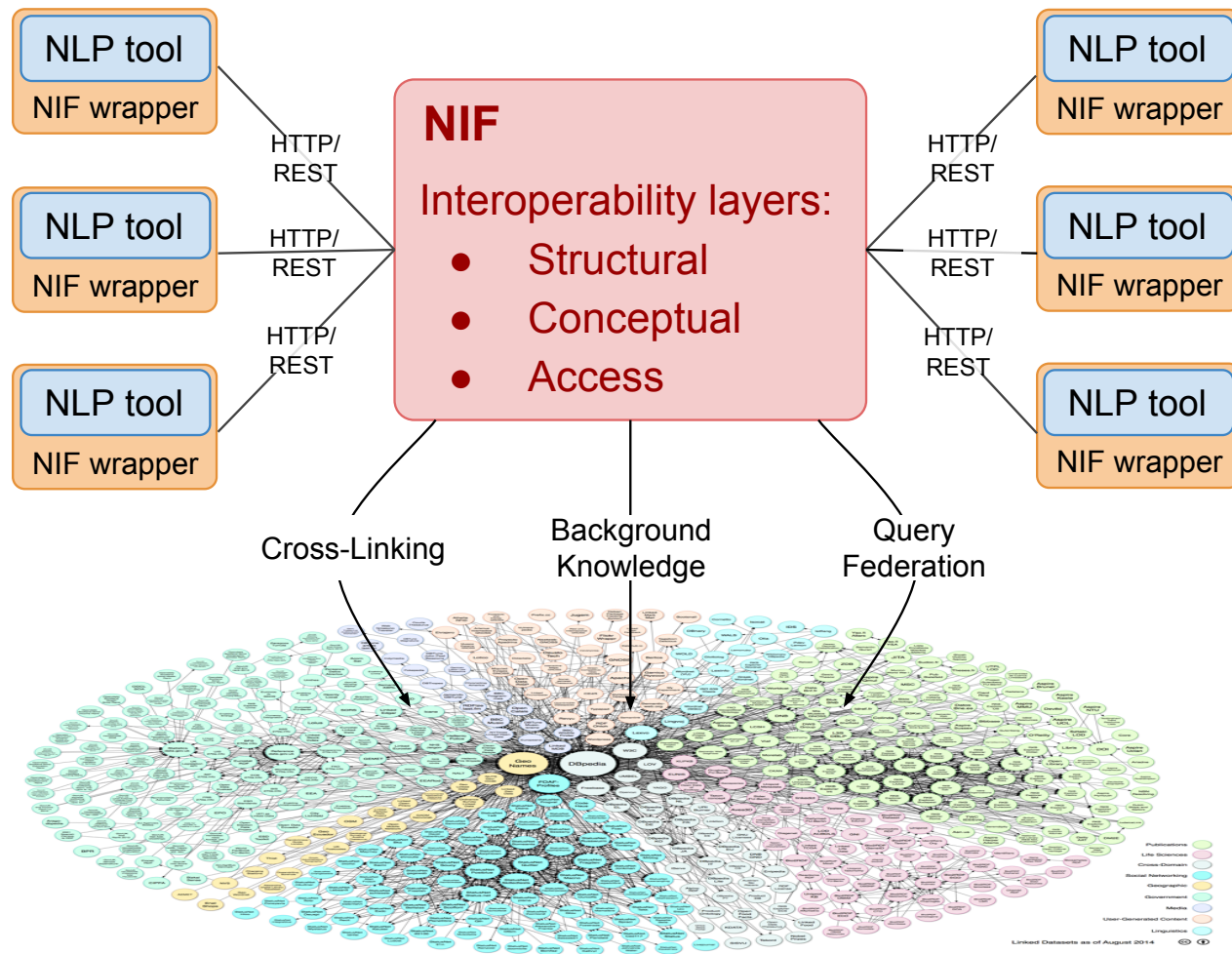
- How to annotate strings
- How to query generated NIF and existing corpora

- Need for integration
 - One-to-one integration
 - Hard to maintain



WTF! Spaghetti ?!!





“My favourite actress is Natalie Portman.”

<#char=3,12>

Tokenizer

```
a nif:String, nif:RFC5147String, nif:Word;
nif:anchorOf      "favourite";
nif:referenceContext <#char=0,>;
nif:beginIndex    "3";
nif:endIndex       "12".
```

<#char=3,12>

Integration through merged RDF

```
a nif:RFC5147String, nif:String;
a nif:Word;
nif:anchorOf      "favourite";
nif:referenceContext <#char=0,>;
nif:beginIndex    "3";
nif:endIndex       "12";
```

@base <http://example.org/prefix>

"My favourite actress is Natalie Portman."

<#char=3,12>

Tokenizer

```
a nif:String, nif:RFC5147String, nif:Word;
nif:anchorOf      "favourite";
nif:referenceContext <#char=0,>;
nif:beginIndex    "3";
nif:endIndex      "12".
```

<#char=3,12>

Snowball Stemmer

```
nif:stem      "favourit".
```

<#char=3,12>

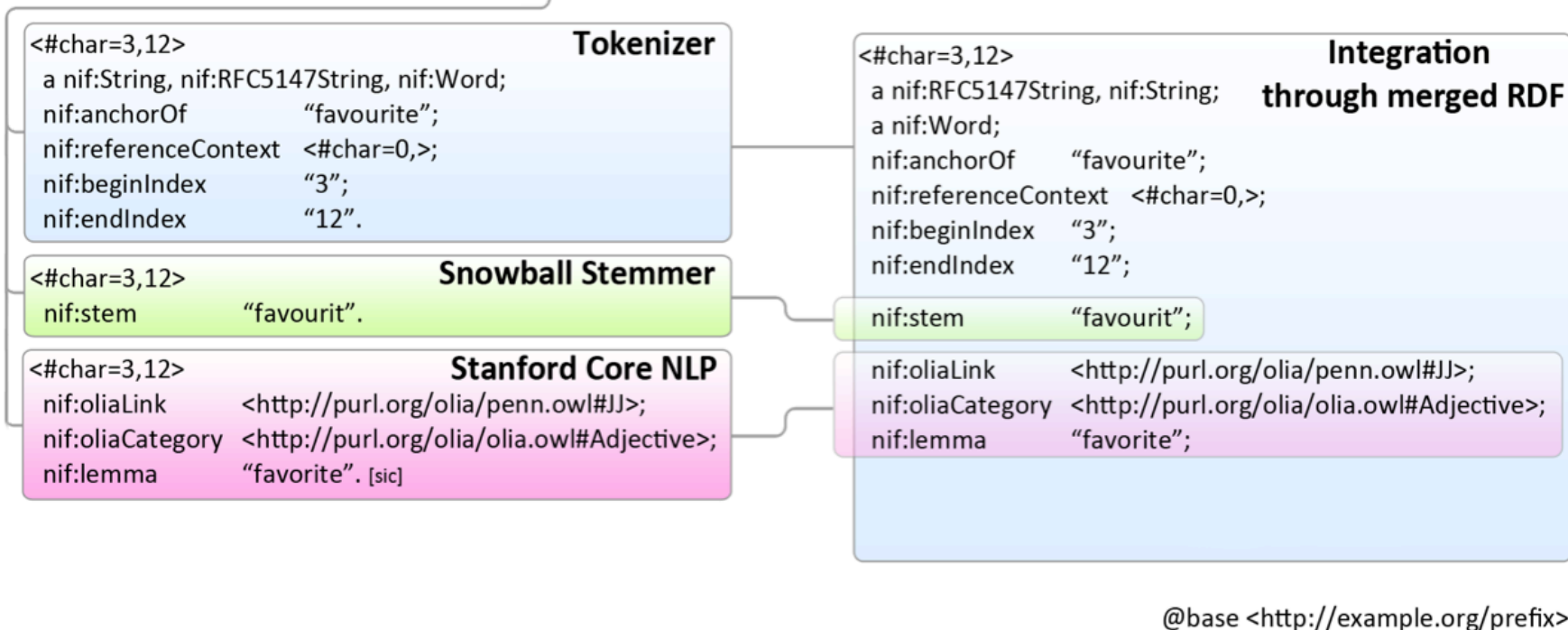
**Integration
through merged RDF**

```
a nif:RFC5147String, nif:String;
a nif:Word;
nif:anchorOf      "favourite";
nif:referenceContext <#char=0,>;
nif:beginIndex    "3";
nif:endIndex      "12";
```

```
nif:stem      "favourit";
```

@base <http://example.org/prefix>

"My favourite actress is Natalie Portman."



“My favourite actress is Natalie Portman.”

Tokenizer

```
<#char=3,12>
a nif:String, nif:RFC5147String, nif:Word;
nif:anchorOf      "favourite";
nif:referenceContext <#char=0,>;
nif:beginIndex    "3";
nif:endIndex      "12".
```

Snowball Stemmer

```
<#char=3,12>
nif:stem      "favourit".
```

Stanford Core NLP

```
<#char=3,12>
nif:oliaLink      <http://purl.org/olia/penn.owl#JJ>;
nif:oliaCategory  <http://purl.org/olia/olia.owl#Adjective>;
nif:lemma         "favorite". [sic]
```

DBpedia Spotlight

```
<#char=3,12>
itsrdf:taldentRef <http://dbpedia.org/resource/Favourite>;
itsrdf:taConfidence "0.10"^^xsd:decimal.
```

Integration through merged RDF

```
<#char=3,12>
a nif:RFC5147String, nif:String;
a nif:Word;
nif:anchorOf      "favourite";
nif:referenceContext <#char=0,>;
nif:beginIndex    "3";
nif:endIndex      "12";

nif:stem      "favourit";

nif:oliaLink      <http://purl.org/olia/penn.owl#JJ>;
nif:oliaCategory  <http://purl.org/olia/olia.owl#Adjective>;
nif:lemma         "favorite";

itsrdf:taldentRef <http://dbpedia.org/resource/Favourite>;
itsrdf:taConfidence "0.10"^^xsd:decimal.
```

@base <http://example.org/prefix>

- Contains hotel reviews and review metadata
- 1760 semi-structured files
- Every file's content becomes nif:Context resource
- Strings addressed with unique URIs


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

<http://example.org/tripadvisor/hotel_100547.dat#char=0,125958>
  a nif:String, nif:RFC5147String, nif:Context;
  nif:isString
  """<Overall Rating>2.5 <Avg. Price>$171 [...]"""^^xsd:string;
  nif:beginIndex "0"^^xsd:nonNegativeInteger ;
  nif:endIndex "125958"^^xsd:nonNegativeInteger ;
  nif:sourceUrl <http://sifaka.cs.uiuc.edu/~wang296/Data/LARA/TripAdvisor/Review_Texts.zip> .
```

- Address the content of the document
- nif:isString contains document content
- In NIF the document != content of the document
- Two documents can have the same content, BUT must not have the same URI

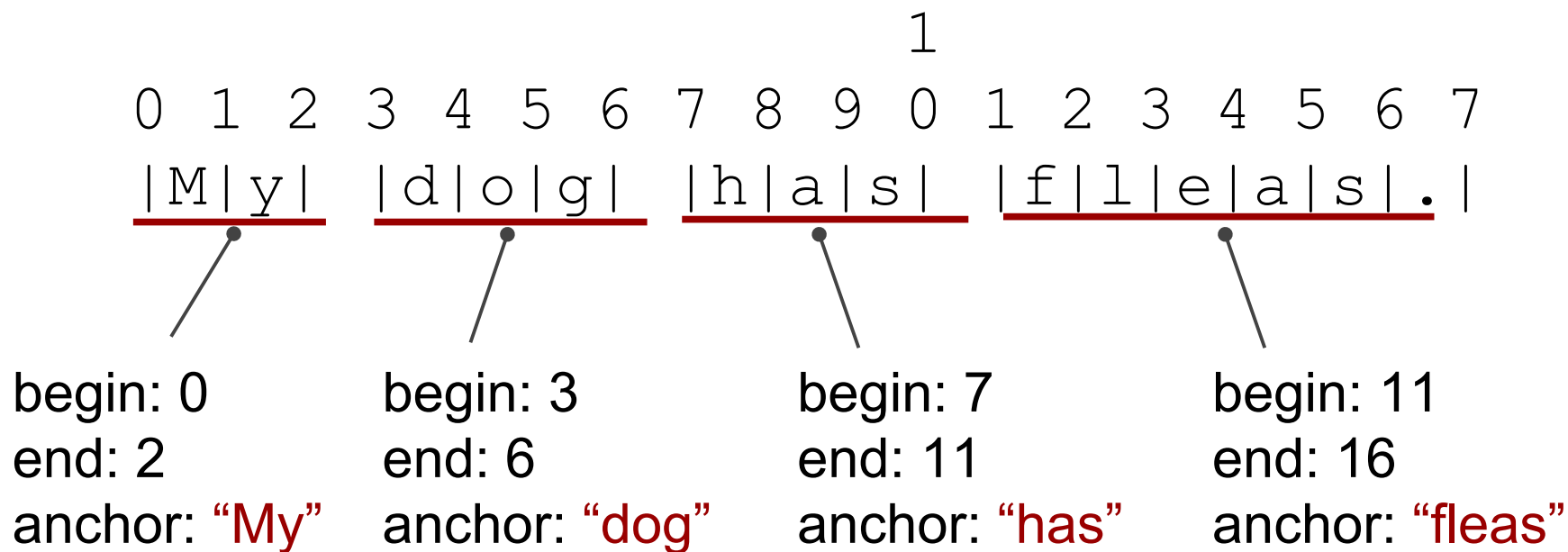
```
<http://example.org/tripadvisor/hotel_100547.dat#char=9010,9529>
  a nif:String, nif:RFC5147String ; a tripadvisor:Rivew
  nif:anchorOf
  """<Author>kekeScotland
<Content>Good Exeprience We got a good deal compared to other Seattle prices and as we were only using the
hotel as a base so we didn't expect too much. Location of the hotel was good and the room was quiet and
clean. We had a safe and free wi-fi. Only downside was the room was on the small side and the bathroom was
tiny.[...] """^^xsd:string;
  nif:beginIndex "9010"^^xsd:nonNetgativeInteger ;
  nif:endIndex "9529"^^xsd:nonNetgativeInteger ;
  nif:referenceContext <http://example.org/tripadvisor/hotel_100547.dat#char=0,125958> .
```

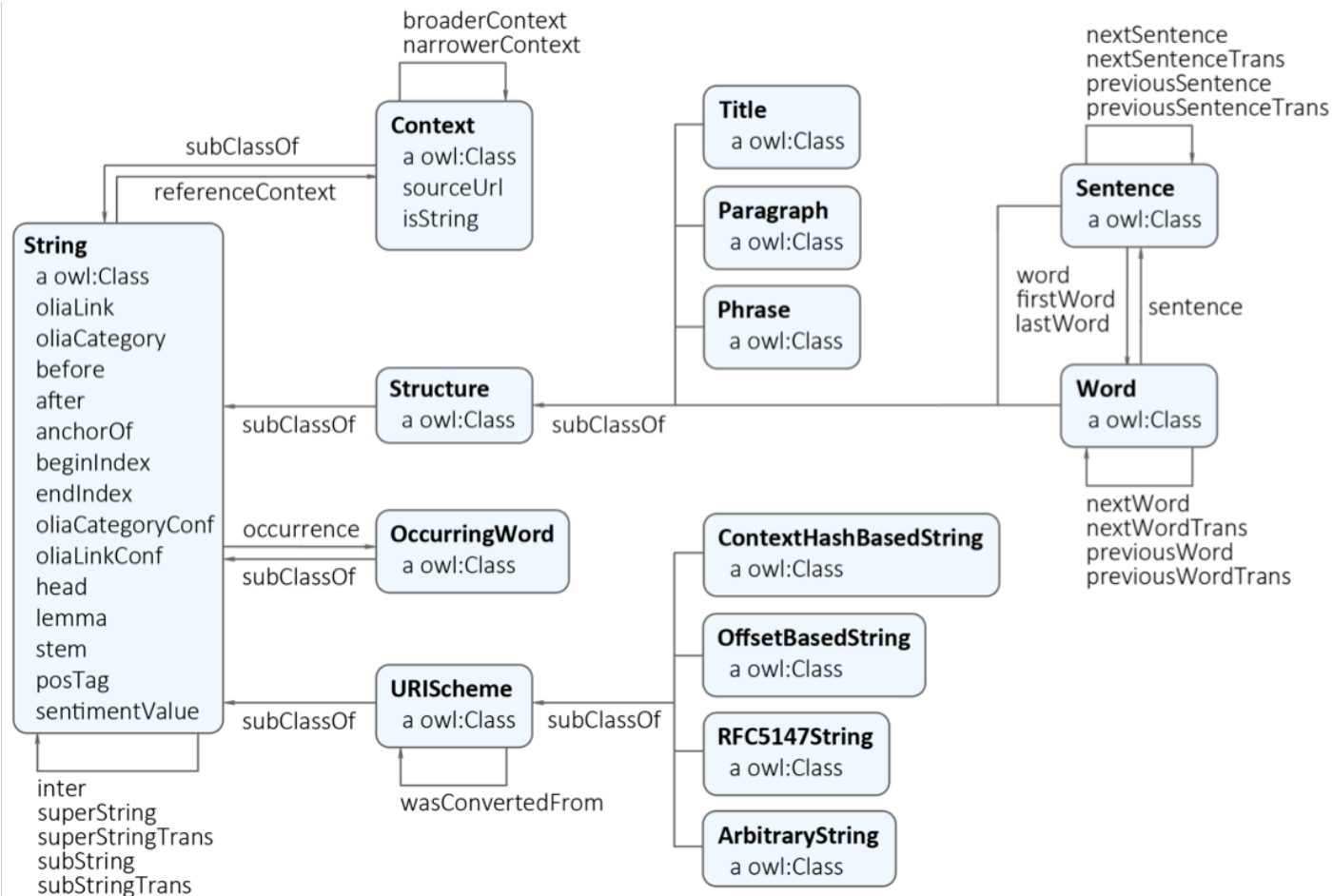
- Address arbitrary strings in the document
- Use string offsets in relation to context to address
- nif:anchorOf contains the string
- Additional properties can be added

```
<http://example.org/tripadvisor/hotel_100547.dat#char=9041,9045>
  a nif:String, nif:RFC5147String, nif:Word;
  nif:anchorOf
  """Good"""^^xsd:string;
  nif:beginIndex "9041"^^xsd:nonNegativeInteger ;
  nif:endIndex "9045"^^xsd:nonNegativeInteger ;
  nif:referenceContext <http://example.org/tripadvisor/hotel_100547.dat#char=0,125958> .

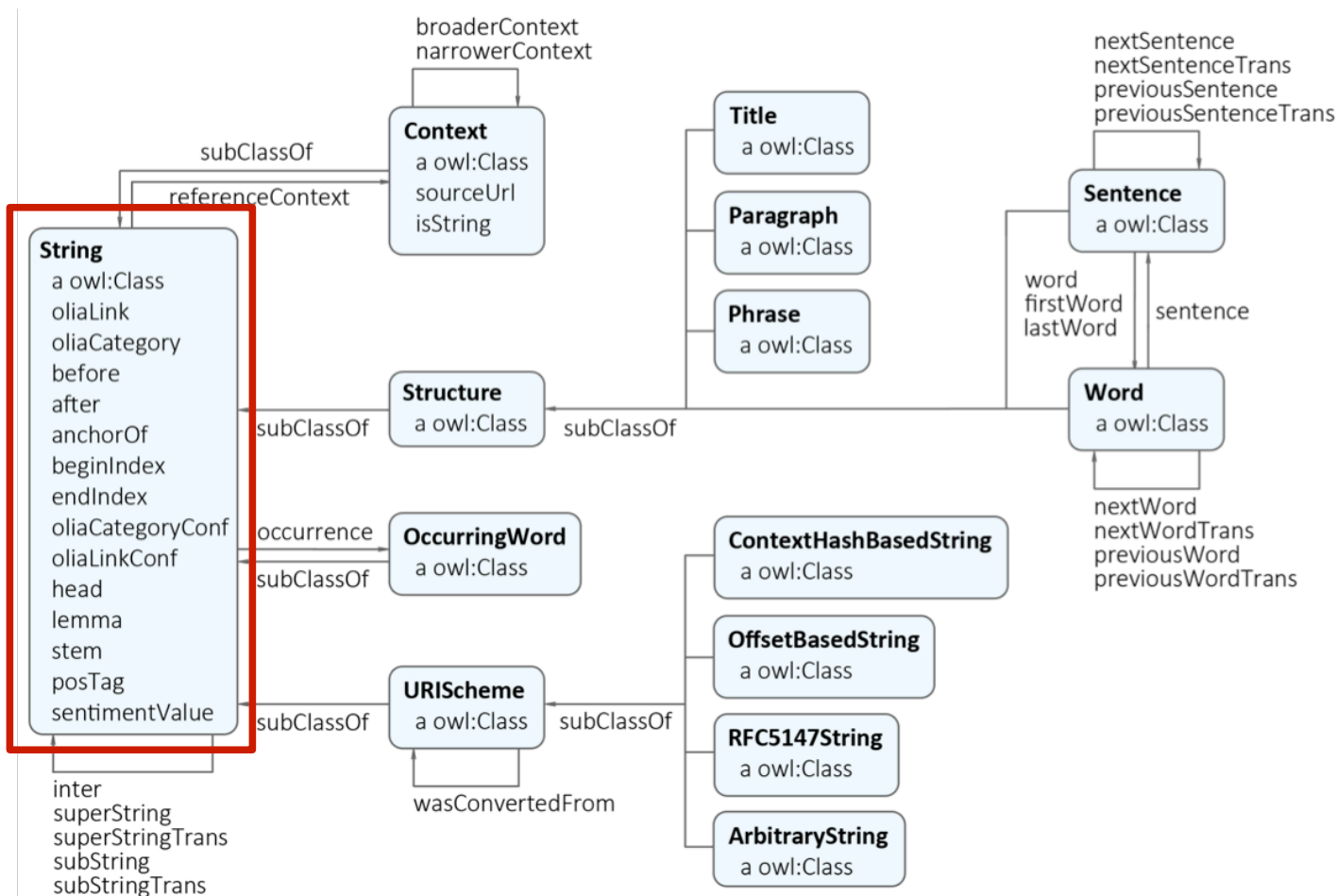
<http://example.org/tripadvisor/hotel_100547.dat#char=9041,9056>
  a nif:String, nif:RFC5147String, nif:Phrase;
  nif:anchorOf
  """Good Exeprience"""^^xsd:string;
  nif:beginIndex "9041"^^xsd:nonNegativeInteger ;
  nif:endIndex "9056"^^xsd:nonNegativeInteger ;
  nif:referenceContext <http://example.org/tripadvisor/hotel_100547.dat#char=0,125958> .
```

- Sentiment values, POS tags and other annotations can be added to the words and phrases

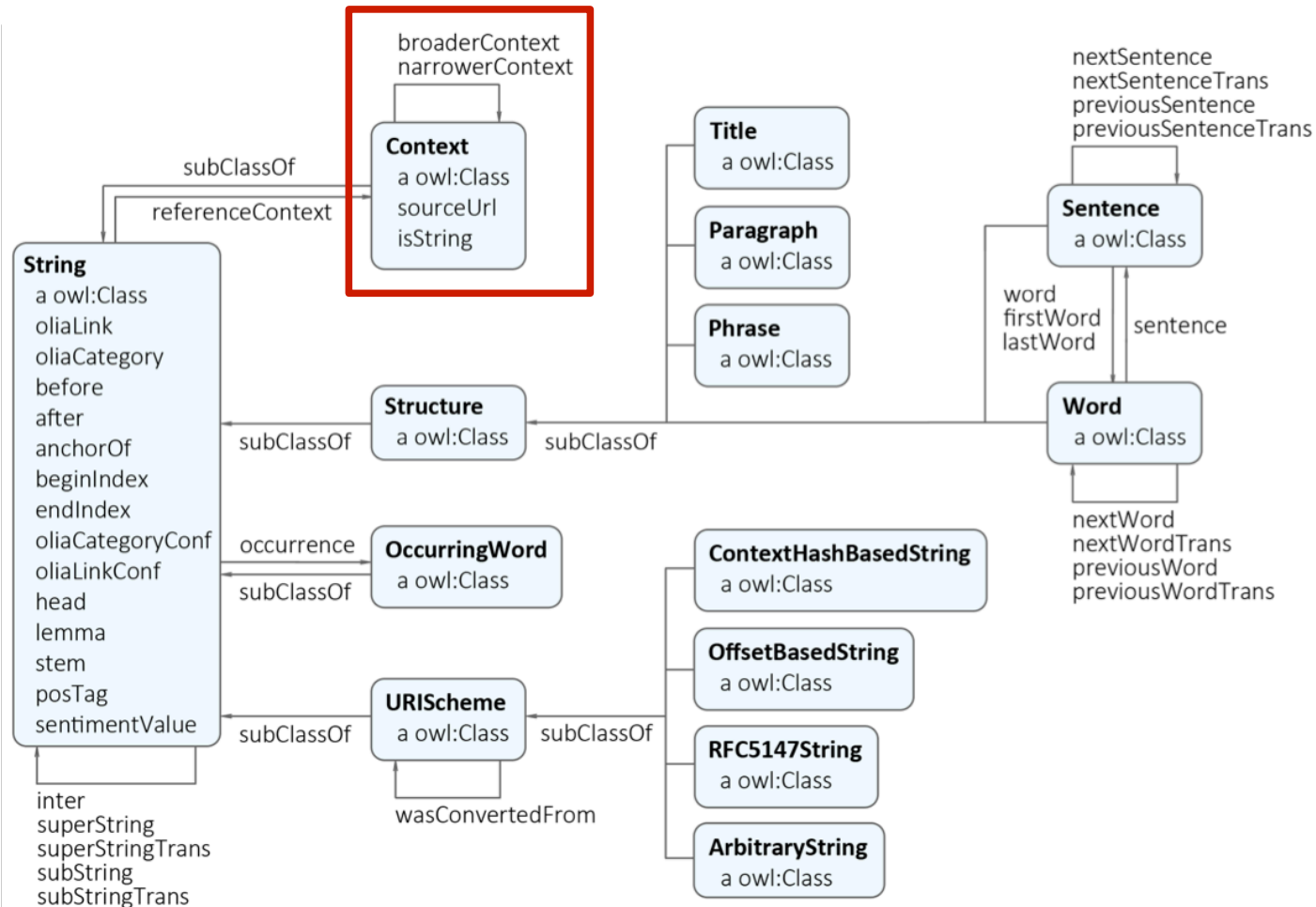




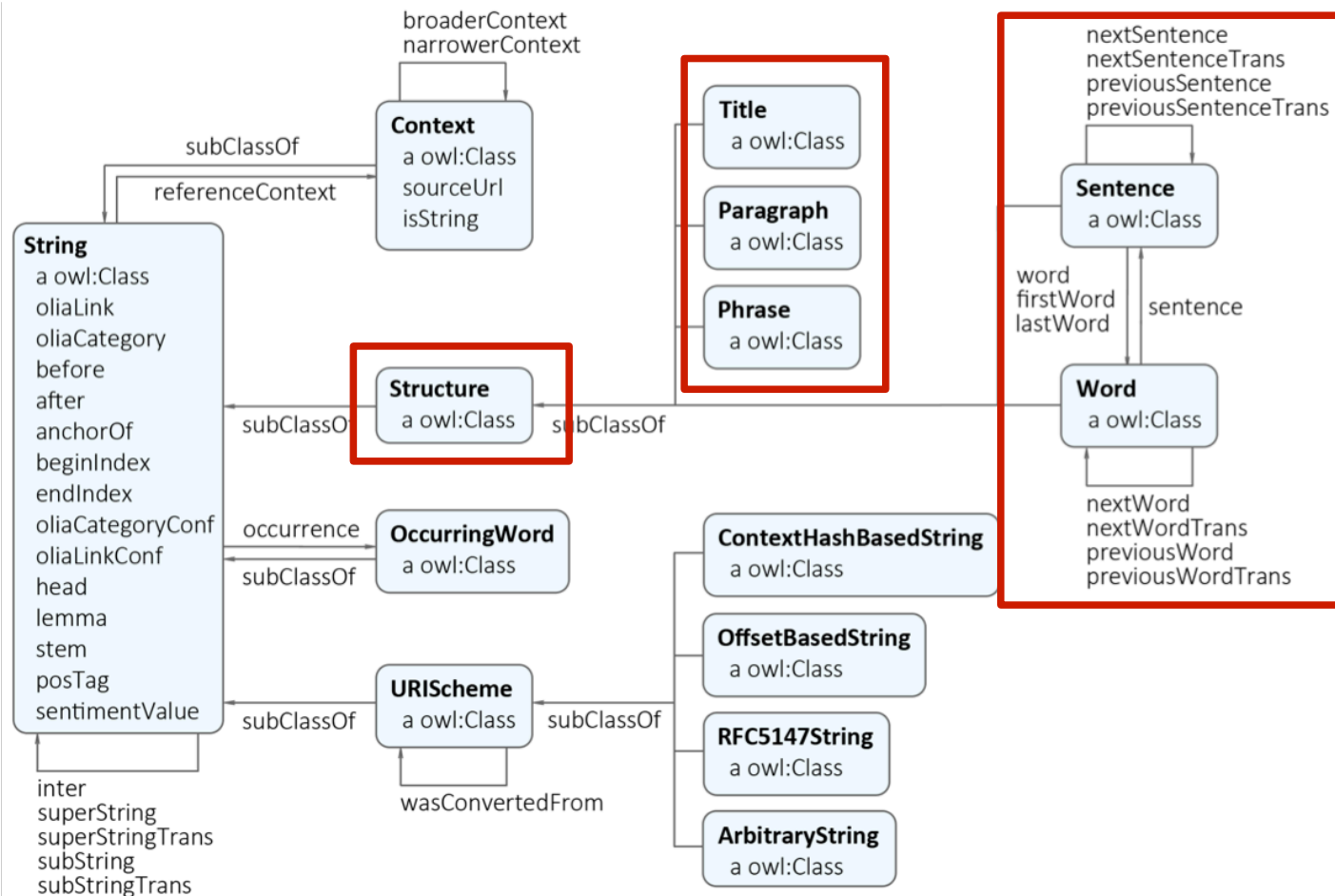
Namespace nif: <<http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#>>



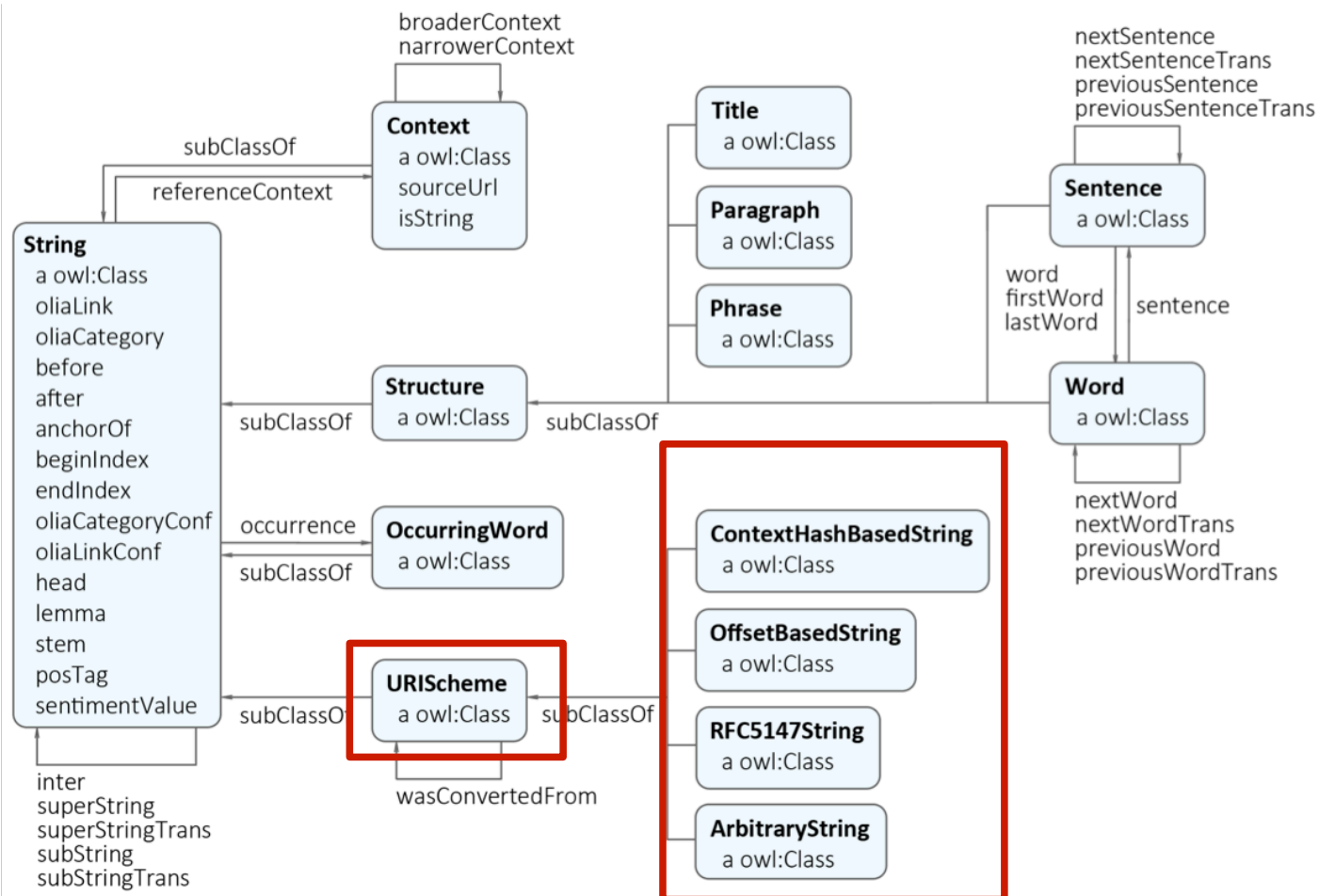
Namespace nif: <<http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#>>



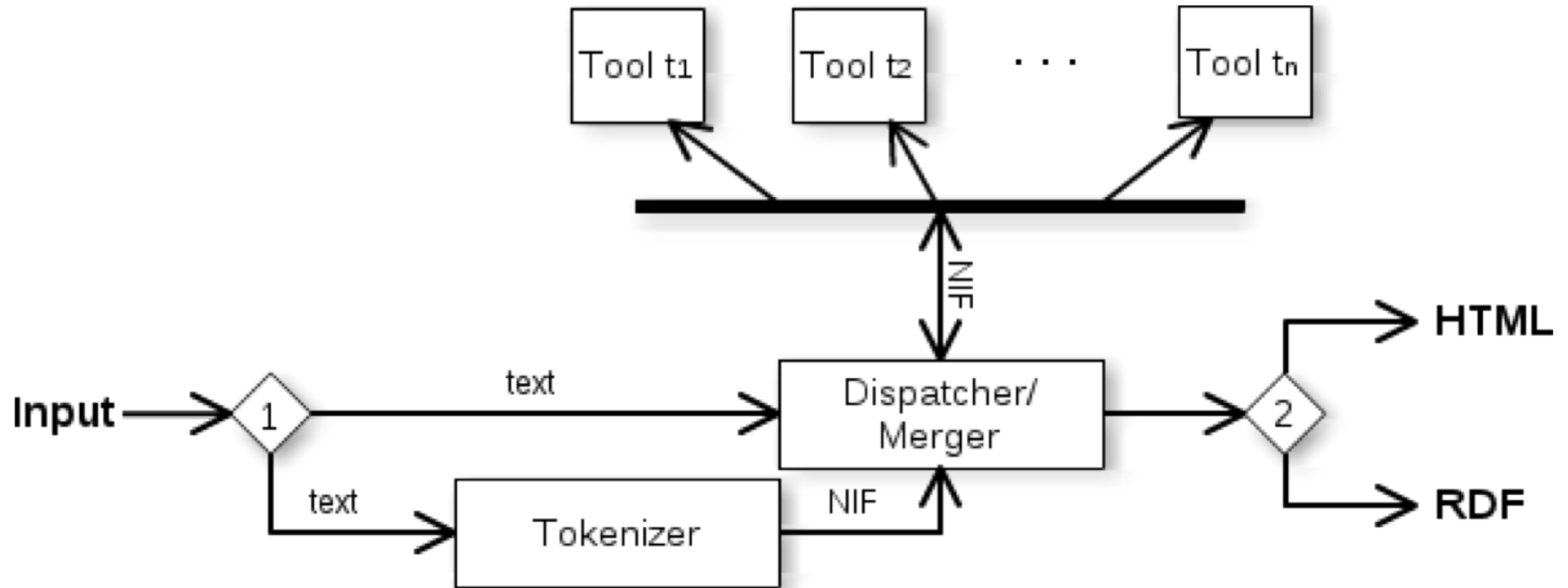
Namespace nif: <<http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#>>



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



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



NIF Combinator: Combining NLP Tool Output

This web site show cases and combines available demo nif web services. Note that the list displayed here is not complete. More tools will be added soon. Read the [full documentation](#) [Bibtex]. The code for this demo is available [here](#). This service also allows you to retrieve the raw data. Use the *format* parameter which supports *rdxml*, *turtle*, *rdjson*, *ntriples* as values. You can also use the *Accept:* header with the appropriate content-type. Note that this service is currently using NIF-2.0 specification.

Action

- ☒ display and merge
- ☐ validate
- ☐ summarize

URI Recipe

- ☒ offset
- ☐ context-hash

Prefix

Tools

- ☒ [Snowball Stemmer](#)
- ☐ [Stanford CoreNLP](#)
- ☐ [OpenNLP](#)
- ☐ [DBpedia Spotlight](#) (Confidence:)

Other NIF service:

Coming soon:

- ☐ [UIMA](#)
- ☐ [Mallet](#)
- ☐ [Gate ANNIE](#)
- ☐ [DBpedia Spotlight](#)

Input text

President Obama on Monday will call for a new minimum tax rate for individuals making more than \$1 million a year.

Merged Output

1. Introduction

- NIF Basics
- **NIF Corpora**
- NIF Tools and Services

2. Hands-on: NIF in action

- How to annotate strings
- How to query generated NIF and existing corpora

Name	Size (in triples)
Wikilinks	500M
News-100	13K
RSS-500	10K
Reuters-128	7K
Spotlight	3K
KORE50	2K
Brown	500K

- Wikipedia abstracts corpus in progress
- Corpora available at <http://datahub.io/dataset?tags=nif&q=nif>
– search for tag “nif” on datahub

- Large scale coreference resolution corpus by Umass/Google
- Over 10M crawled websites that contain text (Named Entities) linked to Wikipedia
- Converted to the NIF format and published as LOD
 - more info here: <http://wiki-link.nlp2rdf.org/>
- Additional processing done to extract relevant text snippets, add DBpedia ontology classes, and coarse-grained classes (entity types)
- Over 500 million triples, 79GB LOD, 12GB gzipped dumps
- Over 30 million links to over 3 million entities

- Converted to the NIF format and published as Linked Data
 - more info here: <http://brown.nlp2rdf.org/>
- Corpus showcases handling of POS tags in NIF
- POS tags mapped via OliA to predefined categories

<#char=643,647>

```
a nif:String , nif:Word , nif:RFC5147String ;
nif:anchorOf "Jury"^^xsd:string ;
nif:referenceContext <#char=0,> ;
nif:oliaLink brown:NN ;
nif:sentence <#char=619,777> ;
nif:beginIndex "643"^^xsd:nonNegativeInteger ;
nif:endIndex "647"^^xsd:nonNegativeInteger .
```

- Categories can be used to query all resources of a certain POS regardless of the tagset used in the corpus

Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

<http://brown.nlp2rdf.org>

Query Text

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

select ?word, ?oliaPos, ?uri where {
  ?uri nif:oliaLink ?oliaPos.
  ?uri nif:anchorOf ?word.
  ?oliaPos rdf:type/rdfs:subClassOf* olia:Noun} Limit 100
```

- Querying all nouns using the OliA mapping

word	oliaPos	uri
"Henry"^^<http://www.w3.org/2001/XMLSchema#string>	http://purl.org/olia/brown.owl#NP	http://brown.nlp2rdf.org/linkeddata.php?t=url&f=xml&i=http://brown.nlp2rdf.org/corpus/a01.xml#char=5755,5760
"Fulton"^^<http://www.w3.org/2001/XMLSchema#string>	http://purl.org/olia/brown.owl#NP	http://brown.nlp2rdf.org/linkeddata.php?t=url&f=xml&i=http://brown.nlp2rdf.org/corpus/a01.xml#char=2767,2773
"W."^^<http://www.w3.org/2001/XMLSchema#string>	http://purl.org/olia/brown.owl#NP	http://brown.nlp2rdf.org/linkeddata.php?t=url&f=xml&i=http://brown.nlp2rdf.org/corpus/a01.xml#char=6510,6512
"Atlanta"^^<http://www.w3.org/2001/XMLSchema#string>	http://purl.org/olia/brown.owl#NP	http://brown.nlp2rdf.org/linkeddata.php?t=url&f=xml&i=http://brown.nlp2rdf.org/corpus/a01.xml#char=955,962
"Bellwood"^^<http://www.w3.org/2001/XMLSchema#string>	http://purl.org/olia/brown.owl#NP	http://brown.nlp2rdf.org/linkeddata.php?t=url&f=xml&i=http://brown.nlp2rdf.org/corpus/a01.xml#char=4977,4985
"Georgia"^^<http://www.w3.org/2001/XMLSchema#string>	http://purl.org/olia/brown.owl#NP	http://brown.nlp2rdf.org/linkeddata.php?t=url&f=xml&i=http://brown.nlp2rdf.org/corpus/a01.xml#char=6206,6213
"Tabb"^^<http://www.w3.org/2001/XMLSchema#string>	http://purl.org/olia/brown.owl#NP	http://brown.nlp2rdf.org/linkeddata.php?t=url&f=xml&i=http://brown.nlp2rdf.org/corpus/a01.xml#char=13059,13063

- Querying all nouns using the OliA mapping

1. Introduction

- NIF Basics
- NIF Corpora
- **NIF Tools and Services**

2. Hands-on: NIF in action

- How to annotate strings
- How to query generated NIF and existing corpora

- Available NIF tools:
 - Stanford Core NLP
 - OpenNLP
 - RDFace
 - Validator
 - CoNLL converter
 - ...

Online REST Webservices		
Service Name	URL	Status
OpenNLP	http://opennlp.nlp2rdf.aksw.org/opennlp?f=text&i=This+is+a+test.&t=direct&modelFolder=model	Online!
Stanford	http://stanford.nlp2rdf.aksw.org/stanfordcorenlp?f=text&i=This+is+a+test.&t=direct	Online!
Snowball	http://snowball.nlp2rdf.aksw.org/snowball?f=text&i=This+is+my+favorite+test.&t=direct	Online!
NIF Validator	http://demo.nlp2rdf.org:9990	Online!
NIF DBpedia Spotlight	http://spotlight.nlp2rdf.aksw.org/spotlight?f=text&i=This+is+Germany.&t=direct&confidence=0.3	Online!
NIF Combinator	http://demo.nlp2rdf.org	Online!

NIF Corpus		
Corpus Name	URL	Size
The LOD Wikilinks corpus	http://wiki-link.nlp2rdf.org/	500M
News-100 corpus	https://raw.githubusercontent.com/AKSW/n3-collection/master/News-100.ttl	13k
RSS-500 corpus	https://raw.githubusercontent.com/AKSW/n3-collection/master/RSS-500.ttl	10k
Reuters-128 corpus	https://raw.githubusercontent.com/AKSW/n3-collection/master/Reuters-128.ttl	7k
Brown corpus	http://brown.nlp2rdf.org/	500k
Add Corpus		

Ontology		NIF Specifications	
NIF Validator TestCase Suite@en	Details	W3C Conversion of NIF to ITS 2.0@en	Details
STC - Simple Test Case Ontology@en	Details	W3C Conversion of ITS 2.0 to NIF@en	Details
RLOG - an RDF Logging Ontology@en	Details	NIF 2.0 Core Specification@en	Details
NIF 2.0 Core Ontology - Validation Model@en	Details	NIF 2.0 Stanbol Profile Specification@en	Details
NIF 2.0 Core Ontology - Inference	Details	NIF 2.0 Stanbol Profile Specification@en	Details

Wikis	
How to publish a txt corpora with NIF as Linked Data@en	Details

Poster	
Promotional poster for Meta Forum@en	Details

Software	
DBpedia Spotlight implementation of NIF@en	Details
NIF 2.0 Validator Tool@en	Details
NIF 2.0 Stanford Core Implementation@en	Details

Demo	
NIF Conversion Demo Service@en	Details
NIF Combinator Demo@en	Details



The interface shows the DBpedia Spotlight logo at the top. Below it, there are two sliders for 'Confidence:' and 'Annotation score:', both set to 0.0. To the right of the sliders is a 'Language:' dropdown menu set to 'English'. Below the sliders is a checkbox for 'n-best candidates' which is unchecked. To the right of the checkbox are two buttons: 'SELECT TYPES...' and 'ANNOTATE'. Below these controls is a text area containing a paragraph about Berlin. The text is: 'First documented in the 13th century, [Berlin](#) was the capital of the Kingdom of [Prussia](#) (1701–1918), the [German Empire](#) (1871–1918), the [Weimar Republic](#) (1919–33) and the [Third Reich](#) (1933–45). [Berlin](#) in the 1920s was the third largest m http://dbpedia.org/resource/Weimar_Republic [War II](#), the city became divided into [East Berlin](#) -- the capital of [East Germany](#) -- and [West Berlin](#), a [West German](#) [exclave](#) surrounded by the [Berlin Wall](#) from 1961–89. Following [German reunification](#) in 1990, the city regained its status as the capital of [Germany](#), hosting 147 foreign embassies.' At the bottom right of the text area is a button labeled 'BACK TO TEXT'.

- <https://github.com/dbpedia-spotlight/dbpedia-spotlight/>

```
@prefix nif: <http://persistence.uni-leipzig.org/nlp2rdf/
  ontologies/nif-core#> .
# POS tags produced by Stanford Parser online demo
# http://nlp.stanford.edu:8080/parser/index.jsp
<http://nlp2rdf.lod2.eu/nif-ws.php#char=13,20>
  rdf:type nif:RFC5147String ;
  nif:anchorOf "actress" ;
  nif:beginIndex "13" ;
  nif:endIndex "20" ;
  nif:referenceContext <http://nlp2rdf.lod2.eu/nif-ws.php#char
    =0,40> ;
# provenance link
nif:oliaProv <http://nlp.stanford.edu:8080/parser/index.jsp> ;
# the tag
nif:oliaLink <http://purl.org/olia/penn.owl#NN> ;
# mappings can be found here: https://github.com/NLP2RDF/
  software
nif:oliaCategory <http://purl.org/olia/olia.owl#Noun> ,
  <http://purl.org/olia/olia.owl#CommonNoun> .
```

```
@prefix nif: <http://persistence.uni-leipzig.org/nlp2rdf/
  ontologies/nif-core#> .
@prefix itsrdf: <http://www.w3.org/2005/11/its/rdf#>.

# POS tags produced by Stanford Parser online demo
# http://nlp.stanford.edu:8080/parser/index.jsp
<http://nlp2rdf.lod2.eu/nif-ws.php#char=24,39>
  rdf:type nif:NamedEntity ;
  nif:anchorOf 'Natalie Portman' ;
  nif:beginIndex '24';
  nif:endIndex '33' ;
  itsrdf:taIdentRef <http://dbpedia.org/resource/Natalie_Portman>
  .
```

```
<http://nlp2rdf.lod2.eu/nif-ws.php#char=24,39>
  itsrdf:taClassRef <http://dbpedia.org/ontology/Actor> ;
  itsrdf:taClassRef <http://dbpedia.org/ontology/Artist> ;
  itsrdf:taClassRef <http://dbpedia.org/ontology/Person> ;
  itsrdf:taClassRef <http://xmlns.com/foaf/0.1/Person> .
```



```
<http://nlp2rdf.lod2.eu/nif-ws.php#char=24,39>
  itsrdf:taIdentRef <http://dbpedia.org/resource/Natalie_Portman
    > ;
  itsrdf:taClassRef <http://dbpedia.org/ontology/Actor> .
# the tool states that the type found is a general property of
  the found entity beyond the context of the current nif:
  Context
<http://dbpedia.org/resource/Natalie_Portman>
  rdf:type <http://dbpedia.org/ontology/Actor> .
```

1. Introduction

- NIF Basics
- NIF Corpora
- NIF Tools and Services

2. Hands-on: NIF in action

- How to annotate strings
- How to query generated NIF and existing corpora

- Github NLP2RDF web page and NIF online demos
 - Dashboard
 - Combinator
- Examples
 - How to annotate string
 - Snowball Steamer, OpenNLP
 - How to query generated NIF and existing corpora

- <https://github.com/NLP2RDF/>

Online REST Webservices		
Service Name	URL	Status
OpenNLP	http://opennlp.nlp2rdf.aksw.org/opennlp?f=text&i=This+is+a+test.&t=direct&modelFolder=model	Online!
Stanford	http://stanford.nlp2rdf.aksw.org/stanfordcorenlp?f=text&i=This+is+a+test.&t=direct	Online!
Snowball	http://snowball.nlp2rdf.aksw.org/snowball?f=text&i=This+is+my+favorite+test.&t=direct	Online!
NIF Validator	http://demo.nlp2rdf.org:9990	Online!
NIF DBpedia Spotlight	http://spotlight.nlp2rdf.aksw.org/spotlight?f=text&i=This+is+Germany.&t=direct&confidence=0.3	Online!
NIF Combinator	http://demo.nlp2rdf.org	Online!

NIF Corpus		
Corpus Name	URL	Size
The LOD Wikilinks corpus	http://wiki-link.nlp2rdf.org/	500M
News-100 corpus	https://raw.githubusercontent.com/AKSW/n3-collection/master/News-100.ttl	13k
RSS-500 corpus	https://raw.githubusercontent.com/AKSW/n3-collection/master/RSS-500.ttl	10k
Reuters-128 corpus	https://raw.githubusercontent.com/AKSW/n3-collection/master/Reuters-128.ttl	7k
Brown corpus	http://brown.nlp2rdf.org/	500k
Add Corpus		

Ontology		NIF Specifications	
NIF Validator TestCase Suite@en	Details	W3C Conversion of NIF to ITS 2.0@en	Details
STC - Simple Test Case Ontology@en	Details	W3C Conversion of ITS 2.0 to NIF@en	Details
RLOG - an RDF Logging Ontology@en	Details	NIF 2.0 Core Specification@en	Details
NIF 2.0 Core Ontology - Validation Model@en	Details	NIF 2.0 Stanbol Profile Specification@en	Details
NIF 2.0 Core Ontology - Inference	Details		

Wikis	
How to publish a txt corpora with NIF as Linked Data@en	Details

Poster	
Promotional poster for Meta Forum@en	Details

Software	
DBpedia Spotlight implementation of NIF@en	Details
NIF 2.0 Validator Tool@en	Details
NIF 2.0 Stanford Core Implementation@en	Details

Demo	
NIF Conversion Demo Service@en	Details
NIF Combinator Demo@en	Details

This web site show cases and combines available demo nif web services. Note that the list displayed here is not complete. More tools will be added soon. Read the [full documentation](#) [Bibtex]. The code for this demo is available [here](#). This service also allows you to retrieve the raw data. Use the *format* parameter which supports *rdxml*, *turtle*, *rdjson*, *ntriples* as values. You can also use the *Accept*: header with the appropriate content-type. Note that this service is currently using NIF-2.0 specification.

Action

☒ display and merge
☐ validate
☐ summarize

URI Recipe

☒ offset
☐ context-hash

Prefix

Tools

☒ [Snowball Stemmer](#)
☐ [Stanford CoreNLP](#)
☐ [OpenNLP](#)
☒ [DBpedia Spotlight](#) (Confidence:)
 Other NIF service:
 Coming soon:
☐ [UIMA](#)
☐ [Mallet](#)
☐ [Gate ANNIE](#)
☐ [DBpedia Spotlight](#)

Try at <http://nlp2rdf.aksw.org>

Input text

President Obama on Monday will call for a new minimum tax rate for individuals making more than \$1 million a year.

Merged Output

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#> .
@prefix rlog: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/rlog#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .

<http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#RFC5147String>
rdf:type owl:Class .

<http://prefix.given.by/theClient#char=16,18> rdf:type nif:Word ,
nif:RFC5147String ;
nif:anchorOf "on" ;

```

Example 1: Snowball Stemmer Wrapper

- Stemming – process for removing suffixes from words
 - CONNECT as common prefix for:
 - CONNECTED
 - CONNECTION
 - CONNECTING
 - CONNECTION

1. Open the USB stick folder
2. Go to “NIF_tutorial_hands_on” folder
3. Open the “instructions.txt” file in a text editor
4. Open a terminal
5. Go to the “jar” folder

6. Copy the first command of the instructions instructions.txt

```
java -jar snowball.jar -f text -i 'My favorite actress is Natalie Portman.'
```

- -f parameter to specify the format
- -i parameter to specify the input text

7. Paste the command in the terminal

```
Milan$ java -jar snowball.jar -f text -i "I am connected"
```

```
Milan$ java -jar snowball.jar -f text -i "I am connected"
```

```
<http://cli.nlp2rdf.org/snowball#char=5,14>
  a nif:Word , nif:RFC5147String ;
  nif:anchorOf "connected" ;
  nif:beginIndex "5" ;
  nif:endIndex "14" ;
  nif:previousWord <http://cli.nlp2rdf.org/snowball#char=2,4> ;
  nif:referenceContext <http://cli.nlp2rdf.org/snowball#char=0,14> ;
  nif:sentence <http://cli.nlp2rdf.org/snowball#char=0,14> ;
  nif:stem "connect" .
```

```
Milan$ java -jar snowball.jar -f text -i "I am connected"
```

Annotation offsets

NIF standard annotations

```
<http://cli.nlp2rdf.org/snowball#char=5,14>
a nif:Word , nif:RFC5147String ;
nif:anchorOf "connected" ;
nif:beginIndex "5" ;
nif:endIndex "14" ;
nif:previousWord <http://cli.nlp2rdf.org/snowball#char=2,4> ;
nif:referenceContext <http://cli.nlp2rdf.org/snowball#char=0,14> ;
nif:sentence <http://cli.nlp2rdf.org/snowball#char=0,14> ;
nif:stem "connect" .
```

Snowball stem annotation

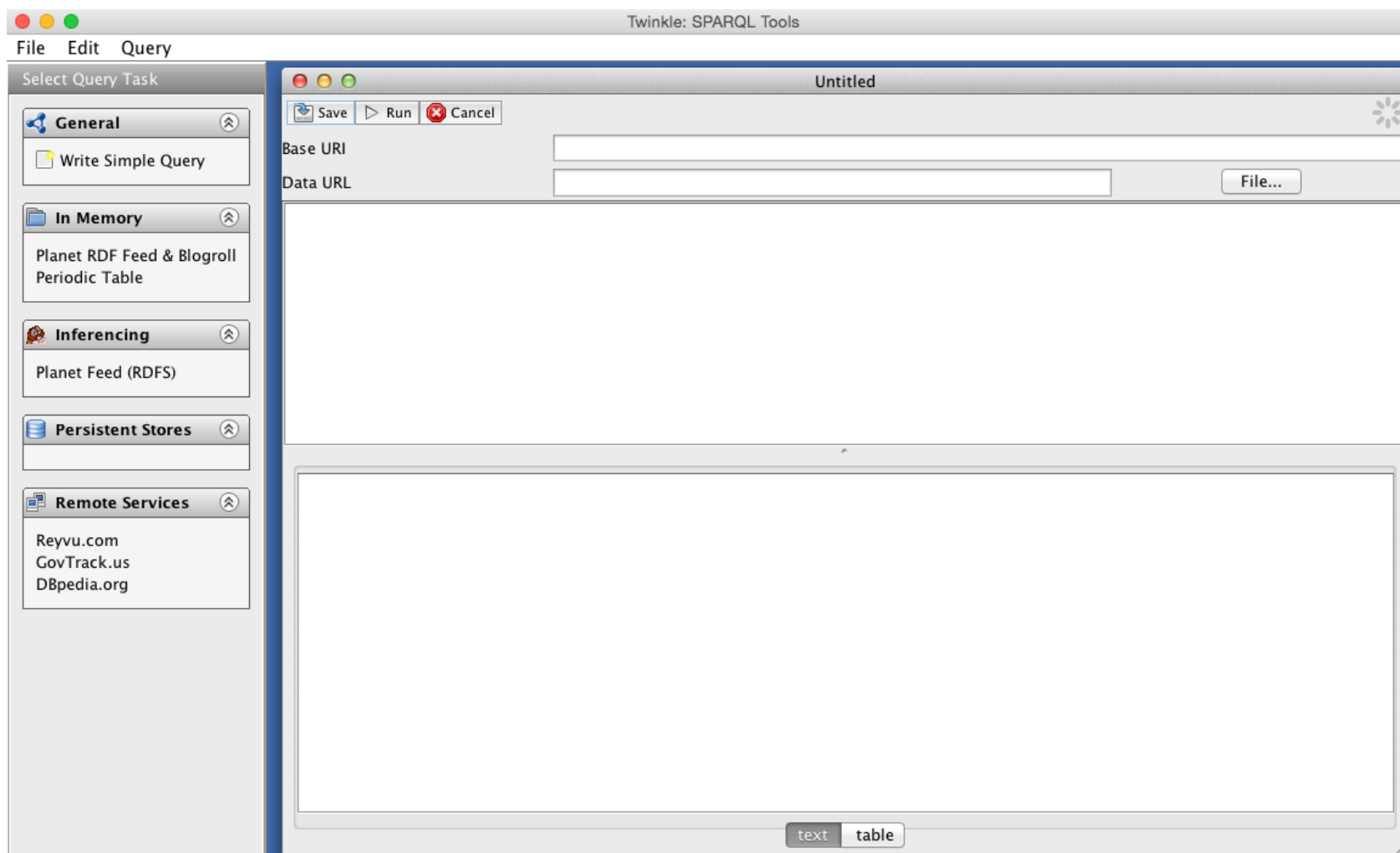
- Back to the terminal and use the second command of the instructions

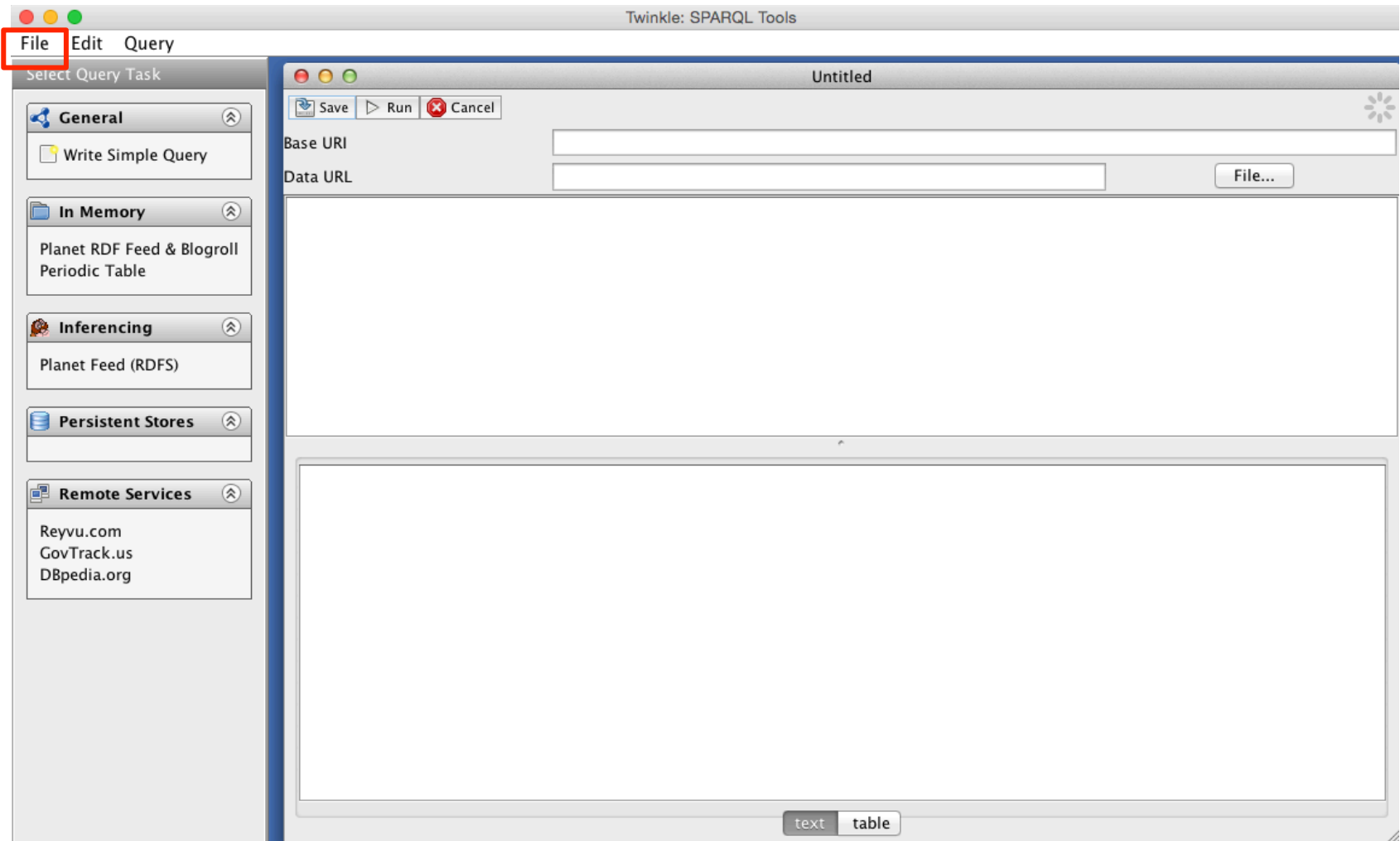
```
java -jar opennlp.jar -f text -i 'My favorite actress is Natalie Portman.' --modelFolder ../model/
```
- The `--modelFolder` parameter set the folder that contains the POS tagging trained models and tokenization
- You might add the parameter `--outfile output.ttl` to store the NIF triples in a file

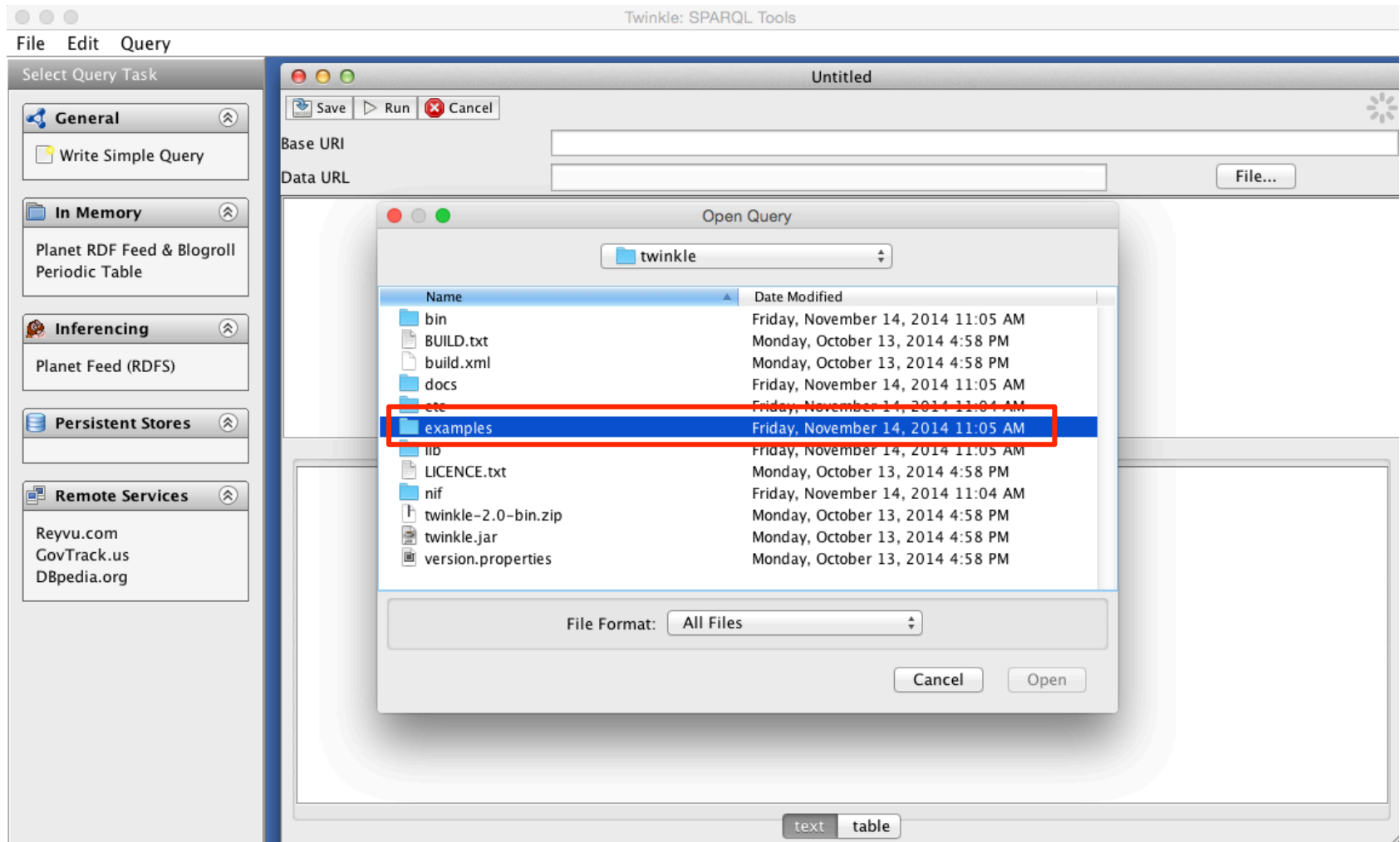
Example 2: Query Brown Corpus

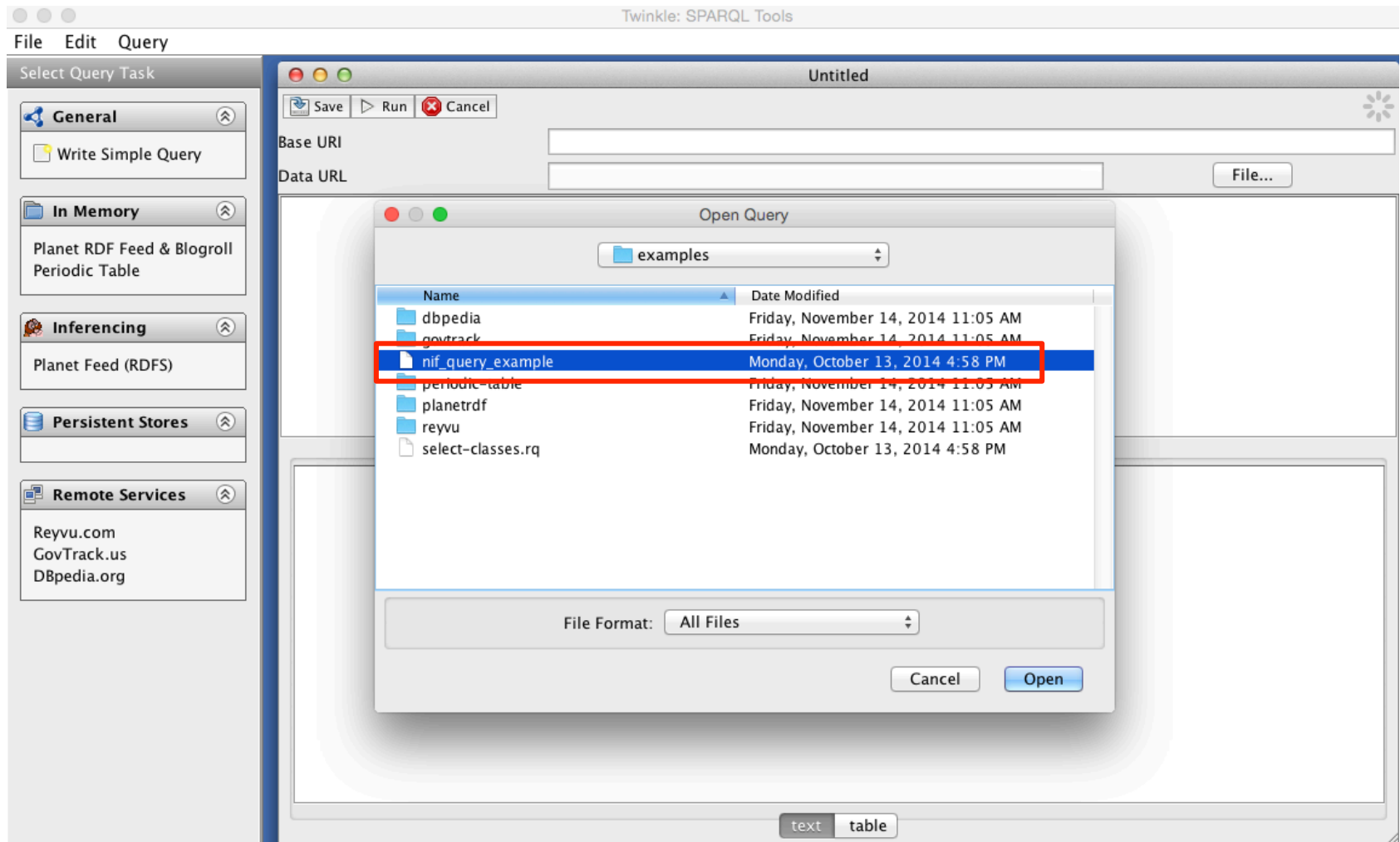
- Open the “/twinkle/example” folder
- Open the NIF_query_example file in a text editor and copy the query
- Open the “/twinkle” folder and run the command

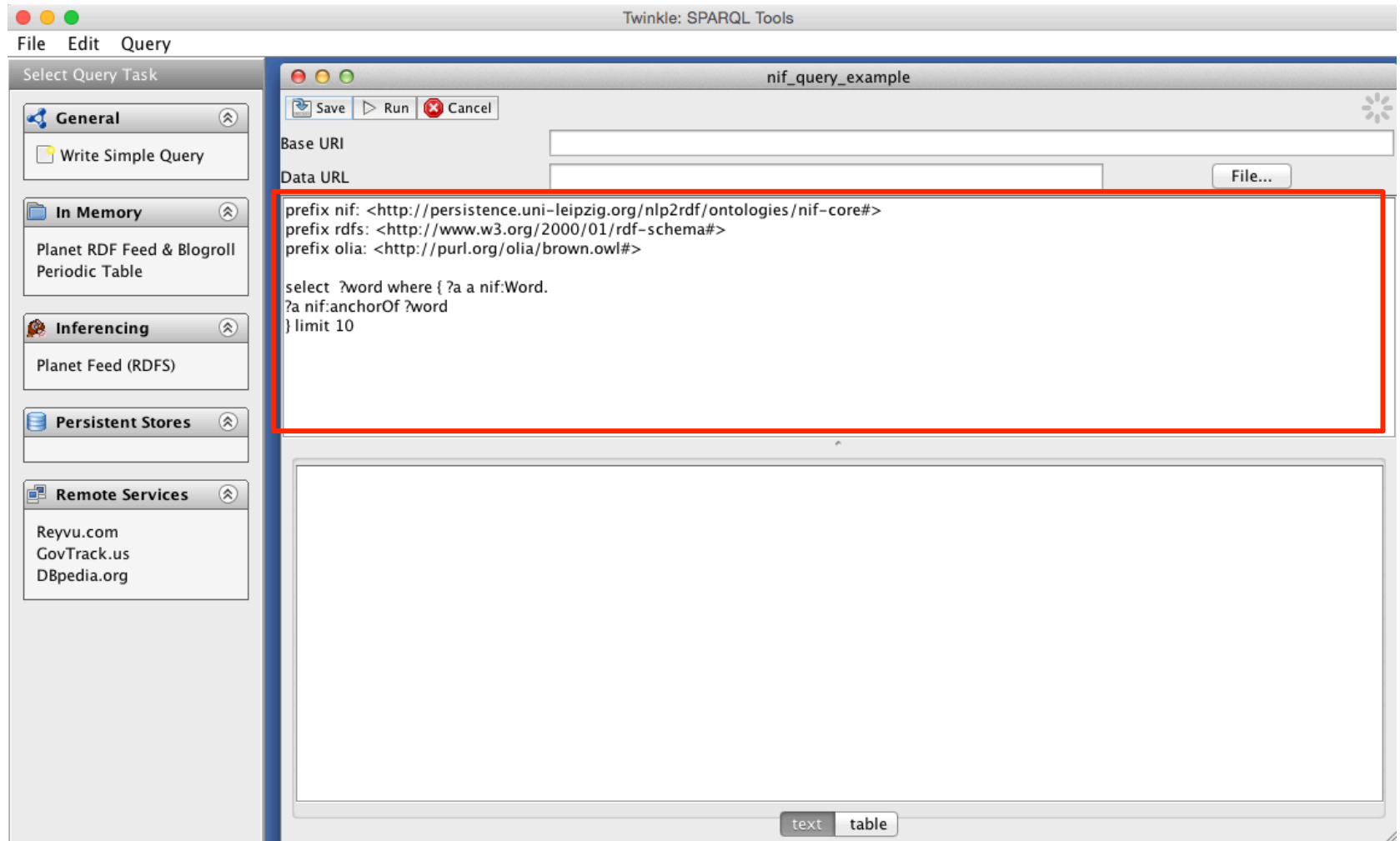
```
java -jar twinkle.jar
```

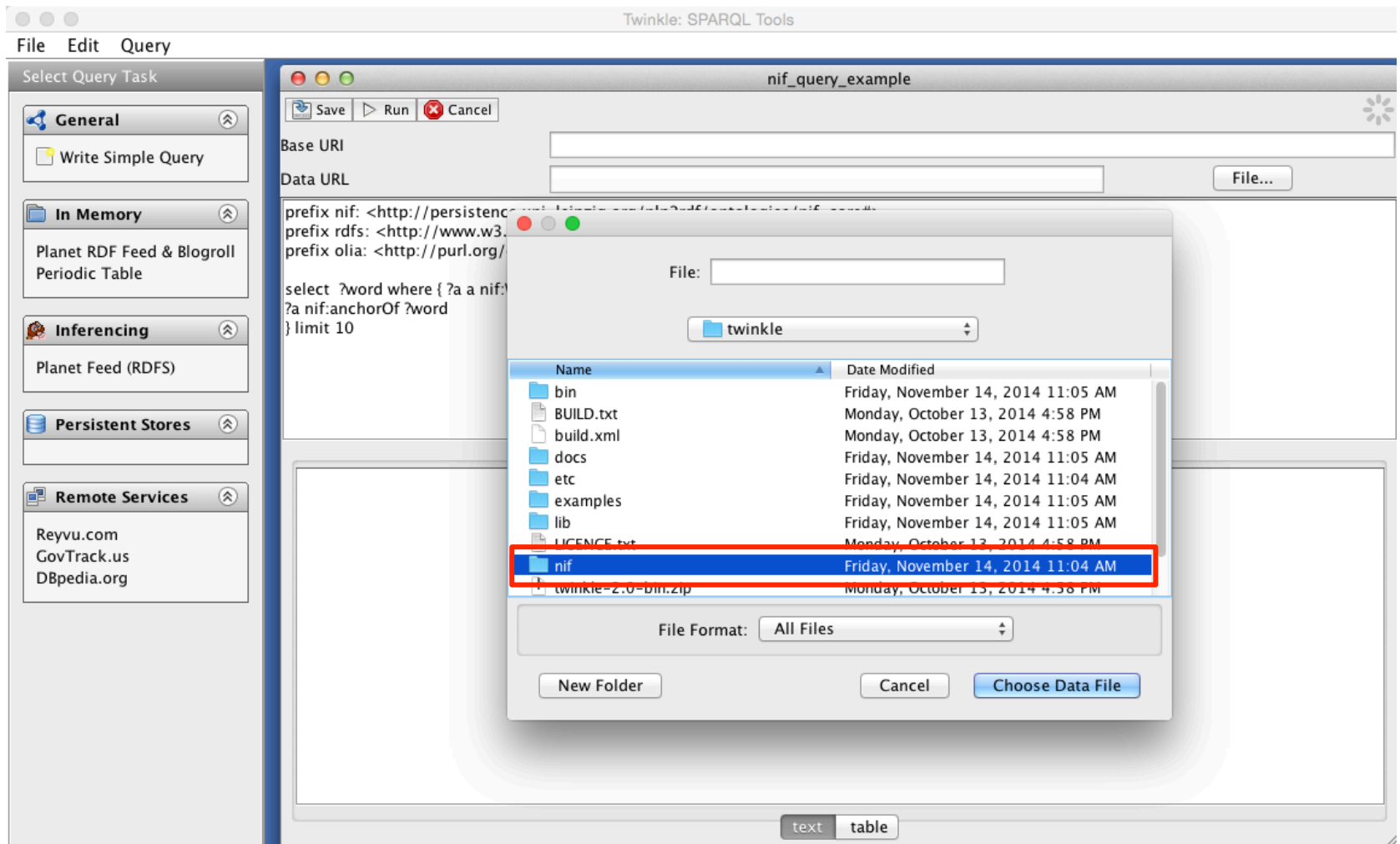


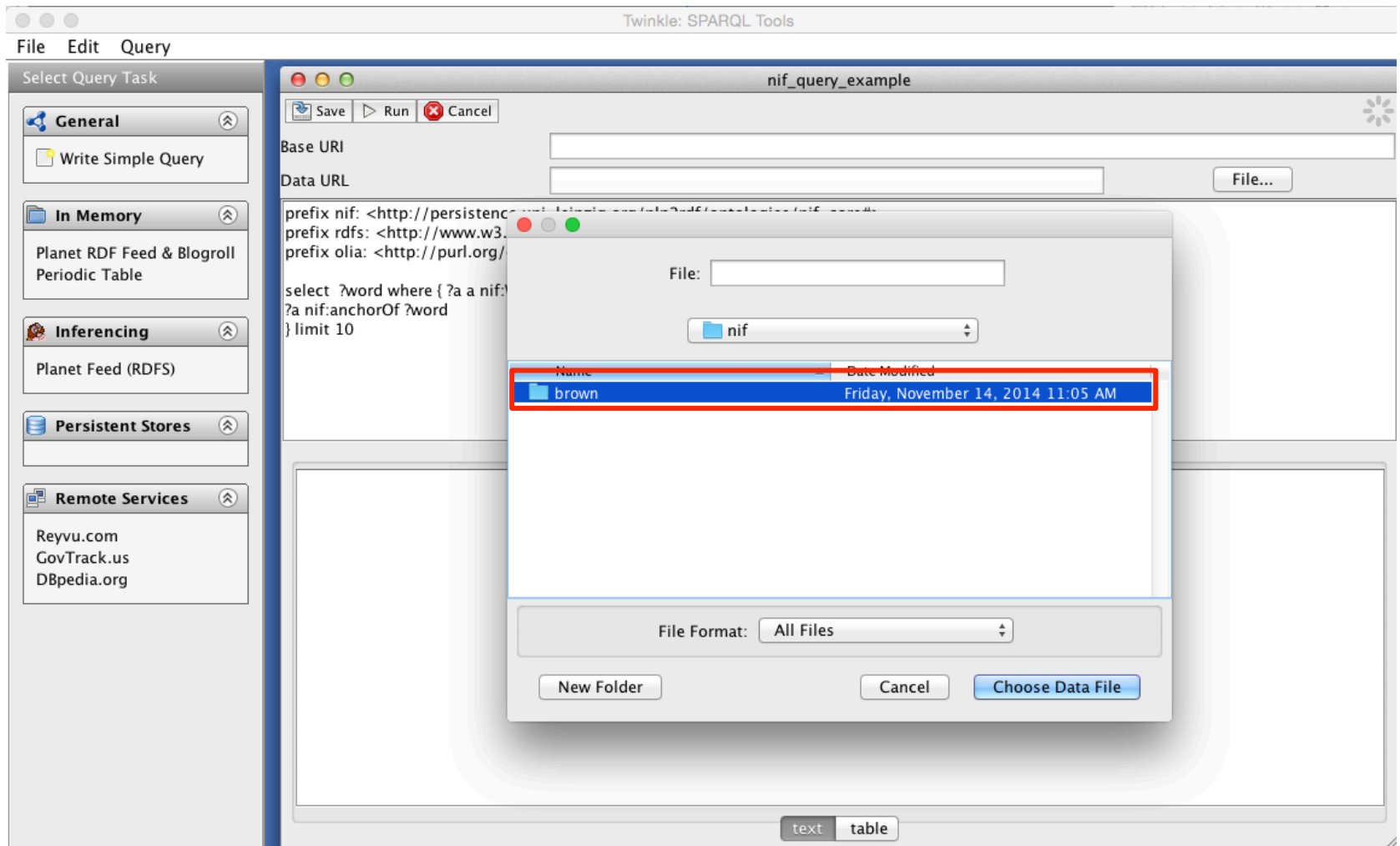


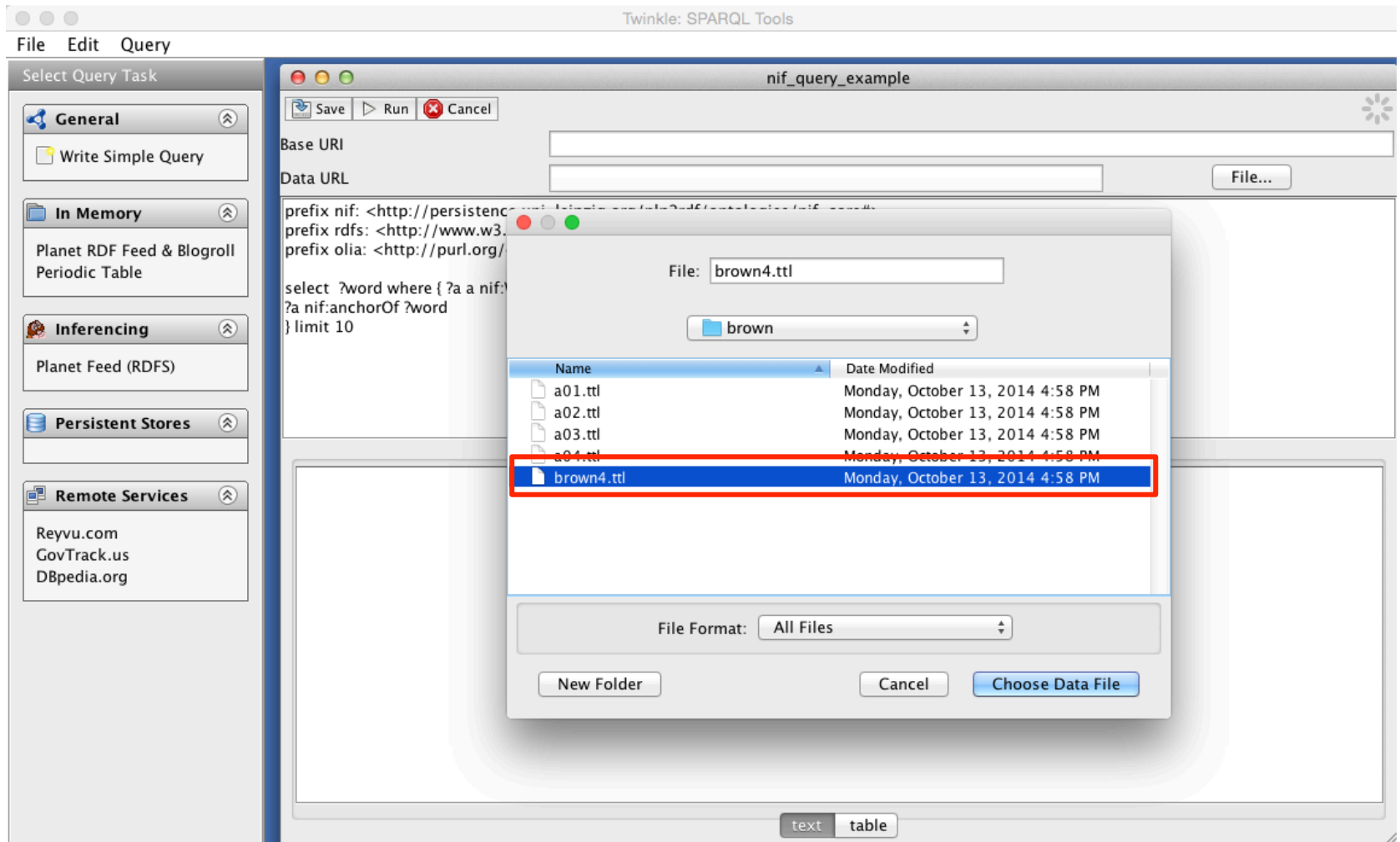


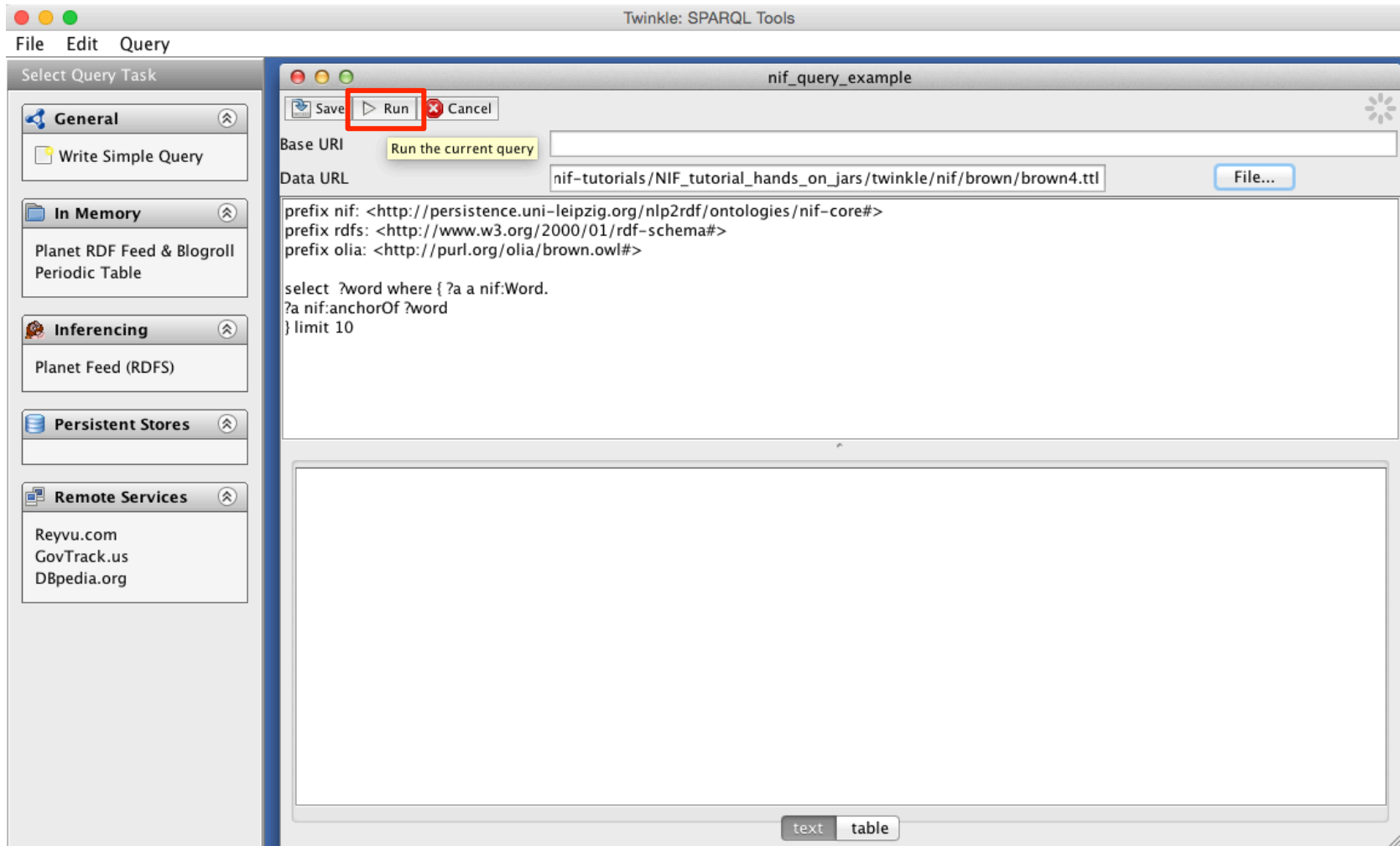


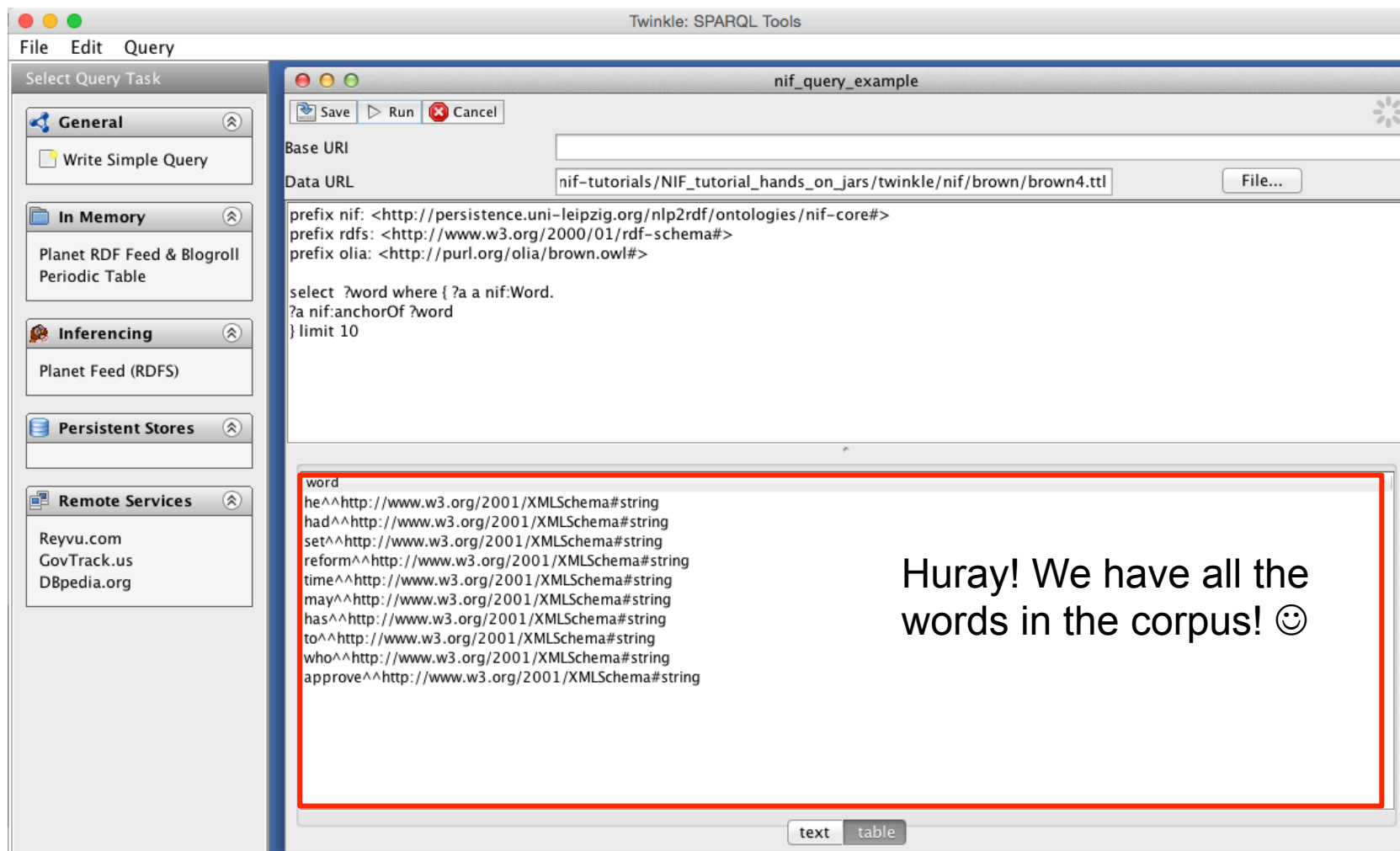












The screenshot shows the Twinkle SPARQL Tools interface. On the left is a sidebar with 'Select Query Task' options: General (Write Simple Query), In Memory (Planet RDF Feed & Blogroll Periodic Table), Inferencing (Planet Feed (RDFS)), Persistent Stores, and Remote Services (Reyvu.com, GovTrack.us, DBpedia.org). The main window, titled 'nif_query_example', contains a query editor with the following content:

```

Base URI
Data URL nif-tutorials/NIF_tutorial_hands_on_jars/twinkle/nif/brown/brown4.ttl
File...

prefix nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix olia: <http://purl.org/olia/brown.owl#>

select ?word where { ?a a nif:Word.
?a nif:anchorOf ?word
} limit 10

```

Below the query editor, a red-bordered box highlights the results, which are displayed as a list of words with their corresponding URIs:

```

word
he^^http://www.w3.org/2001/XMLSchema#string
had^^http://www.w3.org/2001/XMLSchema#string
set^^http://www.w3.org/2001/XMLSchema#string
reform^^http://www.w3.org/2001/XMLSchema#string
time^^http://www.w3.org/2001/XMLSchema#string
may^^http://www.w3.org/2001/XMLSchema#string
has^^http://www.w3.org/2001/XMLSchema#string
to^^http://www.w3.org/2001/XMLSchema#string
who^^http://www.w3.org/2001/XMLSchema#string
approve^^http://www.w3.org/2001/XMLSchema#string

```

To the right of the results, a text overlay reads: "Huray! We have all the words in the corpus! 😊". At the bottom of the results area, there are 'text' and 'table' tabs.

Example 3: Querying your own NIF annotated string

- Querying your own NIF annotated string
 1. Annotated your string using one of the wrappers
 2. Save your annotated sentence to a file
 - set the --outfile parameter
 3. Open Twinkle
 4. Query your string using Twinkle

- Querying your annotated string:
 - nif:Context
 - nif:Sentence
 - nif:anchorOf
 - nif:oliaCategory
 - nif:oliaLink
- ... or practice with the Brown Corpus!

Thank you!

<http://nlp2rdf.org>

<http://github.com/NLP2RDF>