NewsReader: Automatically extracting Events, Entities and Perspectives from Newspapers

Marieke van Erp
marieke.van.erp@vu.nl
http://mariekevanerp.com

VU VRIJE UNIVERSITEIT AMSTERDAM

CLARIAH Common Lab Research Infrastructure for the Arts and Humanities

NewsReader

POST HOC ERGO PROPTER HOC

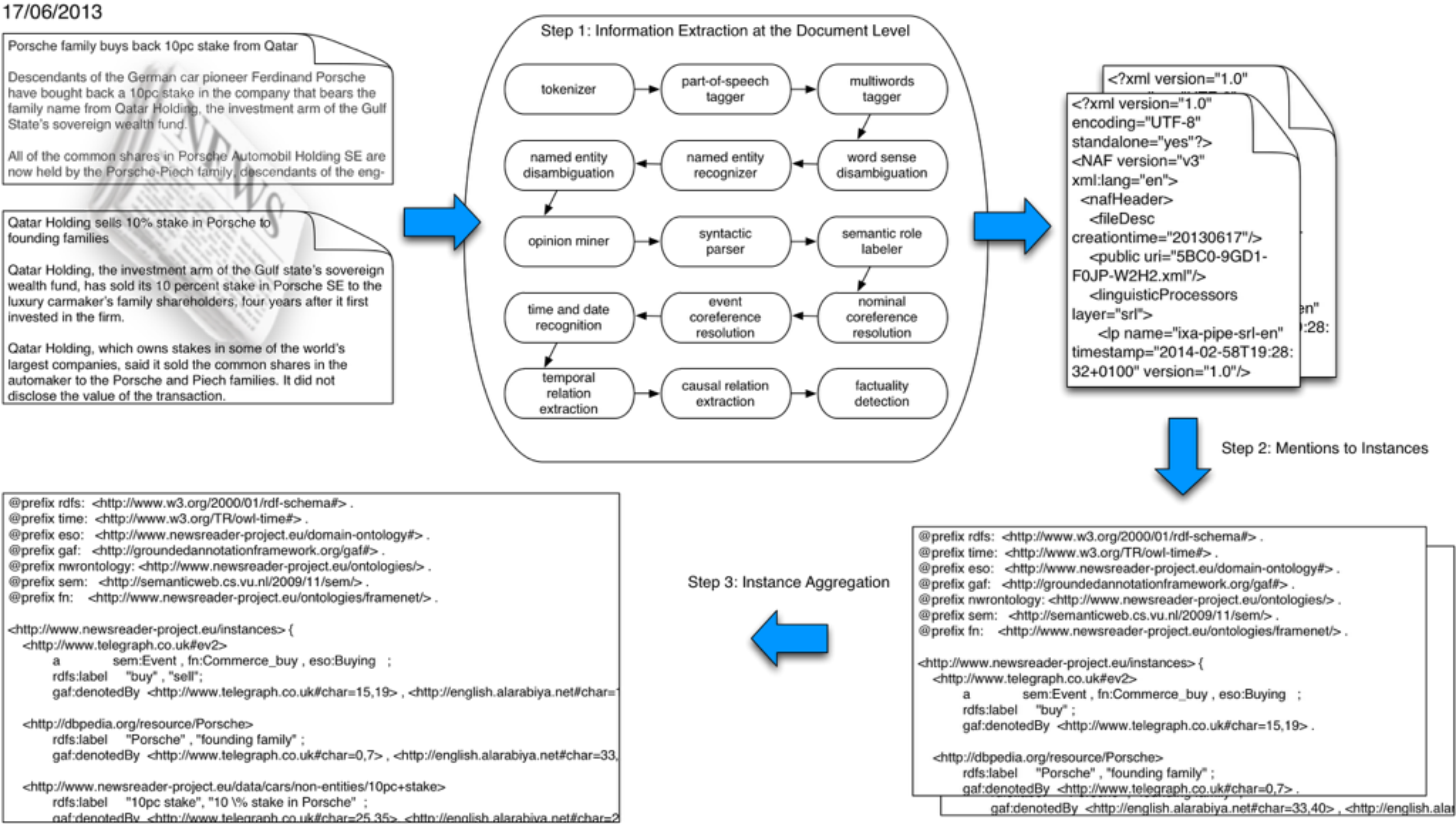
http://www.newsreader-project-eu

- ICT 316404, FP7-ICT-2011-8: Jan. 2013 - Dec. 2015
- Consortium: Vrije Universiteit Amsterdam (NL), The University of The Basque Country (ES), Fondazione Bruno Kessler (IT), LexisNexis (NL), ScraperWiki (now “The Sensible Code Company”, UK) & SynerScope (NL)
- **Read** massive streams of news from many different sources
- **Record** the changes in the world as they are told in the sources in 4 languages: English, Dutch, Spanish and Italian.
- **What** happened, **where** and **when**, **who** was involved.
- From unstructured **Text** to structured **RDF** (through a happy marriage between Computational Linguistics and Semantic Web researchers).
- Who made what statement, where do sources agree and disagree, what is their emotion or judgement: **provenance**

VU VRIJE UNIVERSITEIT AMSTERDAM

CLARIAH
Common Lab Research Infrastructure
for the Arts and Humanities

From Text to RDF



17/06/2013

Porsche family buys back 10pc stake from Qatar

Descendants of the German car pioneer Ferdinand Porsche have bought back a 10pc stake in the company that bears the family name from Qatar Holding, the investment arm of the Gulf State's sovereign wealth fund.

All of the common shares in Porsche Automobil Holding SE are now held by the Porsche-Piech family, descendants of the eng-

Qatar Holding sells 10% stake in Porsche to founding families

Qatar Holding, the investment arm of the Gulf state's sovereign wealth fund, has sold its 10 percent stake in Porsche SE to the luxury carmaker's family shareholders, four years after it first invested in the firm.

Qatar Holding, which owns stakes in some of the world's largest companies, said it sold the common shares in the automaker to the Porsche and Piech families. It did not disclose the value of the transaction.

Step 1: Information Extraction at the Document Level

tokenizer → part-of-speech tagger → multiwords tagger → word sense disambiguation → named entity recognizer → named entity disambiguation → opinion miner → syntactic parser → semantic role labeler → nominal coreference resolution → event coreference resolution → time and date recognition → temporal relation extraction → causal relation extraction → factuality detection

```
<?xml version="1.0"
```

```
<?xml version="1.0"
encoding="UTF-8"
standalone="yes"?>
<NAF version="v3"
xml:lang="en">
  <nafHeader>
    <fileDesc
creationtime="20130617"/>
    <public uri="5BC0-9GD1-
F0JP-W2H2.xml"/>
    <linguisticProcessors
layer="srl">
      <lp name="ixa-pipe-srl-en"
timestamp="2014-02-58T19:28:
32+0100" version="1.0"/>
```

Step 2: Mentions to Instances

```
@prefix rdfs:  <http://www.w3.org/2000/01/rdf-schema#> .
@prefix time:  <http://www.w3.org/TR/owl-time#> .
@prefix eso:   <http://www.newsreader-project.eu/domain-ontology#> .
@prefix gaf:   <http://groundedannotationframework.org/gaf#> .
@prefix nwrontology: <http://www.newsreader-project.eu/ontologies/> .
@prefix sem:   <http://semanticweb.cs.vu.nl/2009/11/sem/> .
@prefix fn:    <http://www.newsreader-project.eu/ontologies/framenet/> .

<http://www.newsreader-project.eu/instances> {
  <http://www.telegraph.co.uk#ev2>
        a             sem:Event , fn:Commerce_buy , eso:Buying   ;
        rdfs:label     "buy" ;
        gaf:denotedBy  <http://www.telegraph.co.uk#char=15,19> .

  <http://dbpedia.org/resource/Porsche>
        rdfs:label     "Porsche" , "founding family" ;
        gaf:denotedBy  <http://www.telegraph.co.uk#char=0,7> .
```

```
        gaf:denotedBy  <http://english.alarabiya.net#char=33,40> , <http://english.alar
```

Step 3: Instance Aggregation

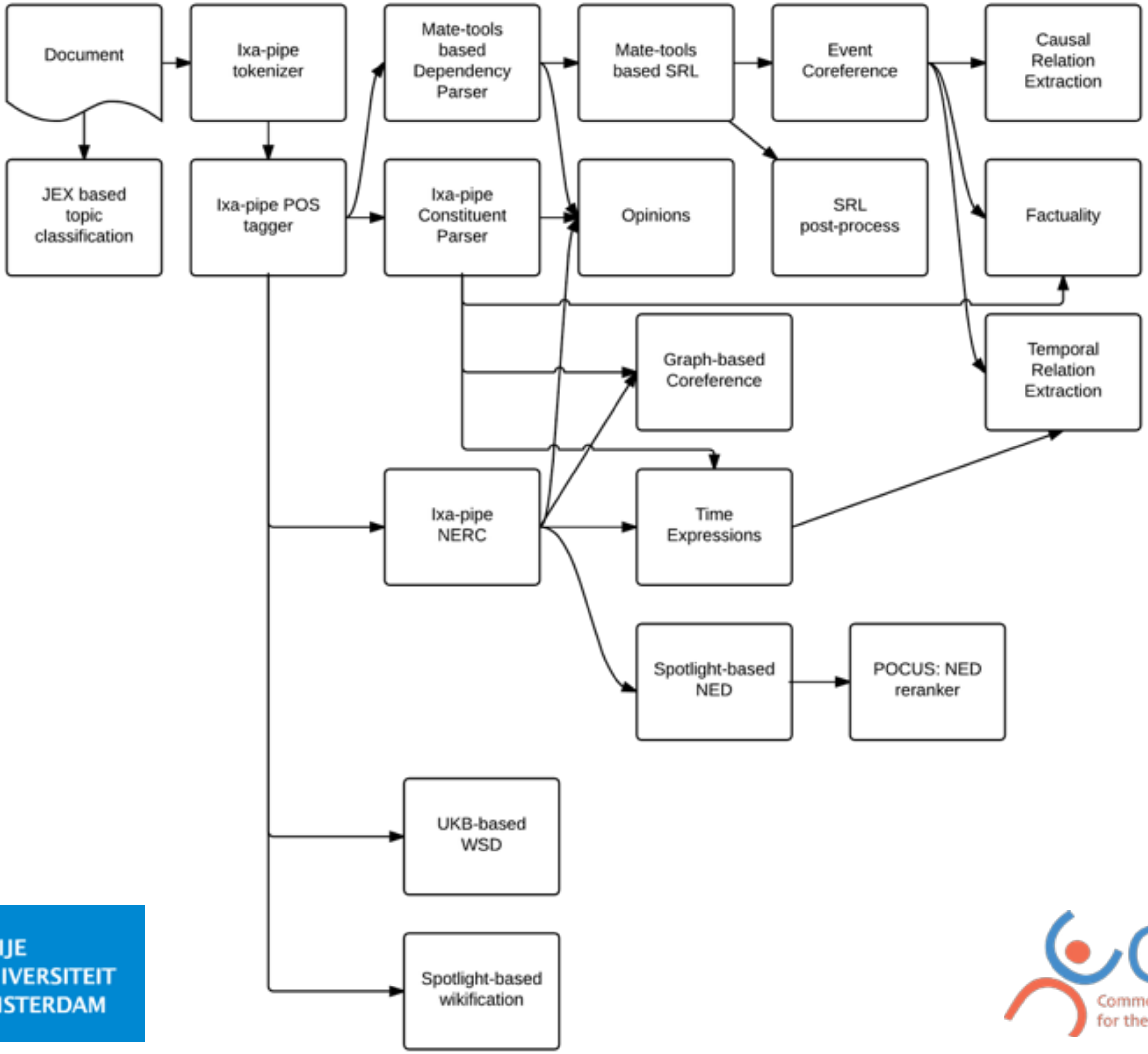
```
@prefix rdfs:  <http://www.w3.org/2000/01/rdf-schema#> .
@prefix time:  <http://www.w3.org/TR/owl-time#> .
@prefix eso:   <http://www.newsreader-project.eu/domain-ontology#> .
@prefix gaf:   <http://groundedannotationframework.org/gaf#> .
@prefix nwrontology: <http://www.newsreader-project.eu/ontologies/> .
@prefix sem:   <http://semanticweb.cs.vu.nl/2009/11/sem/> .
@prefix fn:    <http://www.newsreader-project.eu/ontologies/framenet/> .

<http://www.newsreader-project.eu/instances> {
  <http://www.telegraph.co.uk#ev2>
        a             sem:Event , fn:Commerce_buy , eso:Buying   ;
        rdfs:label     "buy" , "sell";
        gaf:denotedBy  <http://www.telegraph.co.uk#char=15,19> , <http://english.alarabiya.net#char=1

  <http://dbpedia.org/resource/Porsche>
        rdfs:label     "Porsche" , "founding family" ;
        gaf:denotedBy  <http://www.telegraph.co.uk#char=0,7> , <http://english.alarabiya.net#char=33,

  <http://www.newsreader-project.eu/data/cars/non-entities/10pc+stake>
        rdfs:label     "10pc stake", "10 \% stake in Porsche"  ;
        gaf:denotedBy  <http://www.telegraph.co.uk#char=25,35>, <http://english.alarabiya.net#char=2
```

Natural Language Processing Pipeline

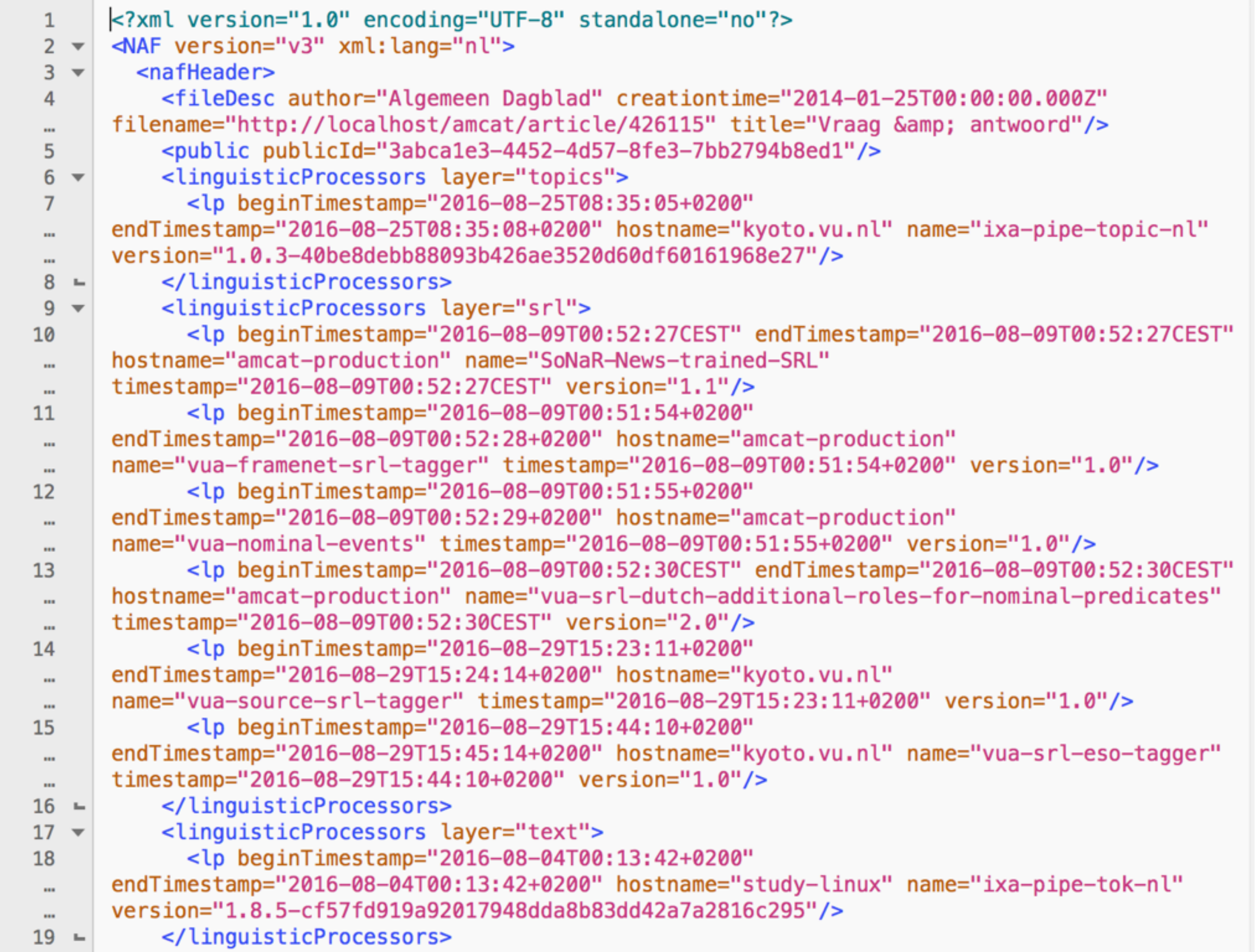


VRIJE UNIVERSITEIT AMSTERDAM

CLARIAH Common Lab Research Infrastructure for the Arts and Humanities

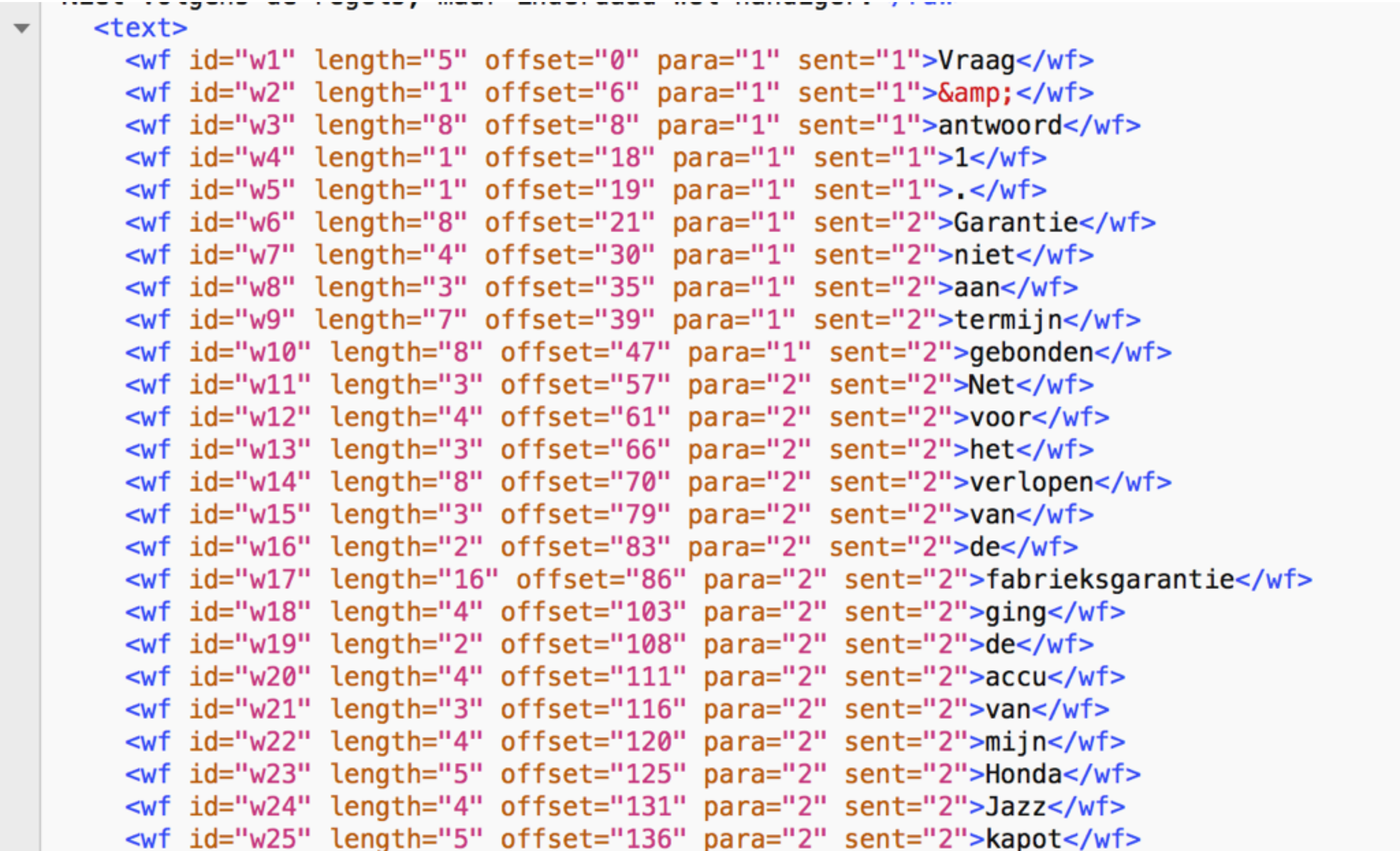
NLP Annotation Format

- Stand-off XML
- Based on KAF, TAF, LAF and uses URIs (from RDF)
- NAF-FoLiA converters are in progress
- Each annotation receives a new layer



```xml
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<NAF version="v3" xml:lang="nl">
  <nafHeader>
    <fileDesc author="Algemeen Dagblad" creationtime="2014-01-25T00:00:00.000Z" filename="http://localhost/amcat/article/426115" title="Vraag &amp; antwoord"/>
    <public publicId="3abca1e3-4452-4d57-8fe3-7bb2794b8ed1"/>
    <linguisticProcessors layer="topics">
      <lp beginTimestamp="2016-08-25T08:35:05+0200" endTimestamp="2016-08-25T08:35:08+0200" hostname="kyoto.vu.nl" name="ixa-pipe-topic-nl" version="1.0.3-40be8debb88093b426ae3520d60df60161968e27"/>
    </linguisticProcessors>
    <linguisticProcessors layer="srl">
      <lp beginTimestamp="2016-08-09T00:52:27CEST" endTimestamp="2016-08-09T00:52:27CEST" hostname="amcat-production" name="SoNaR-News-trained-SRL" timestamp="2016-08-09T00:52:27CEST" version="1.1"/>
      <lp beginTimestamp="2016-08-09T00:51:54+0200" endTimestamp="2016-08-09T00:52:28+0200" hostname="amcat-production" name="vua-framenet-srl-tagger" timestamp="2016-08-09T00:51:54+0200" version="1.0"/>
      <lp beginTimestamp="2016-08-09T00:51:55+0200" endTimestamp="2016-08-09T00:52:29+0200" hostname="amcat-production" name="vua-nominal-events" timestamp="2016-08-09T00:51:55+0200" version="1.0"/>
      <lp beginTimestamp="2016-08-09T00:52:30CEST" endTimestamp="2016-08-09T00:52:30CEST" hostname="amcat-production" name="vua-srl-dutch-additional-roles-for-nominal-predicates" timestamp="2016-08-09T00:52:30CEST" version="2.0"/>
      <lp beginTimestamp="2016-08-29T15:23:11+0200" endTimestamp="2016-08-29T15:24:14+0200" hostname="kyoto.vu.nl" name="vua-source-srl-tagger" timestamp="2016-08-29T15:23:11+0200" version="1.0"/>
      <lp beginTimestamp="2016-08-29T15:44:10+0200" endTimestamp="2016-08-29T15:45:14+0200" hostname="kyoto.vu.nl" name="vua-srl-eso-tagger" timestamp="2016-08-29T15:44:10+0200" version="1.0"/>
    </linguisticProcessors>
    <linguisticProcessors layer="text">
      <lp beginTimestamp="2016-08-04T00:13:42+0200" endTimestamp="2016-08-04T00:13:42+0200" hostname="study-linux" name="ixa-pipe-tok-nl" version="1.8.5-cf57fd919a92017948dda8b83dd42a7a2816c295"/>
    </linguisticProcessors>
```

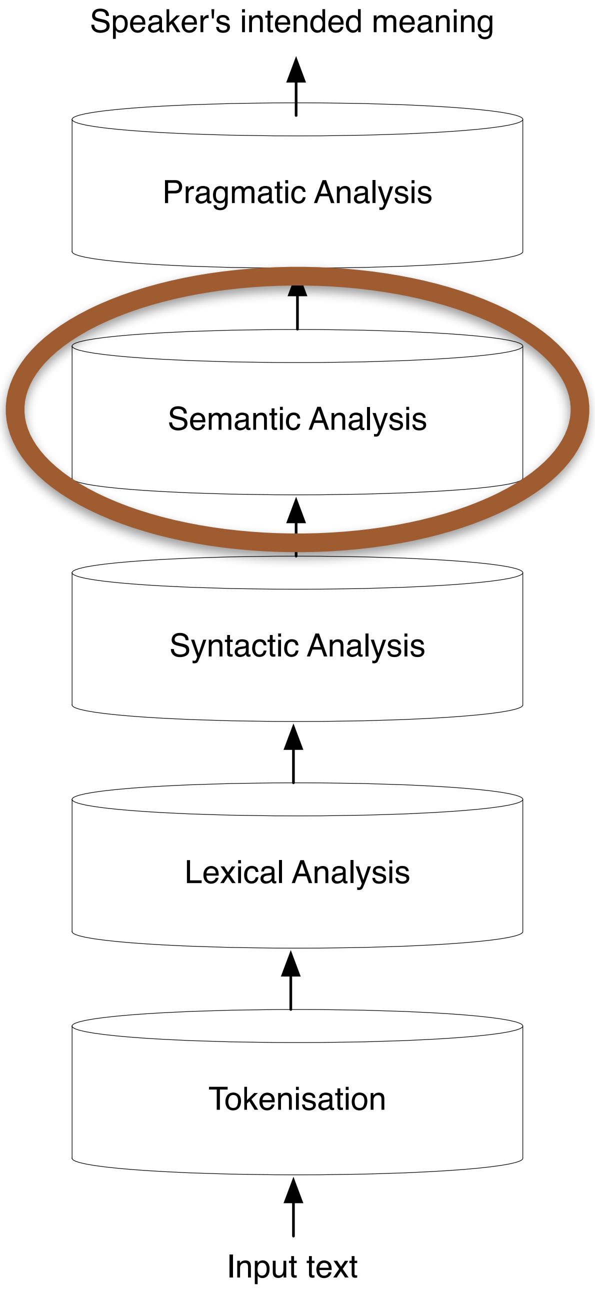
NLP Annotation Format



```
niet volgens de regels, maar inderdaad niet handig?.</raw>
  <text>
    <wf id="w1" length="5" offset="0" para="1" sent="1">Vraag</wf>
    <wf id="w2" length="1" offset="6" para="1" sent="1">&amp;</wf>
    <wf id="w3" length="8" offset="8" para="1" sent="1">antwoord</wf>
    <wf id="w4" length="1" offset="18" para="1" sent="1">1</wf>
    <wf id="w5" length="1" offset="19" para="1" sent="1">.</wf>
    <wf id="w6" length="8" offset="21" para="1" sent="2">Garantie</wf>
    <wf id="w7" length="4" offset="30" para="1" sent="2">niet</wf>
    <wf id="w8" length="3" offset="35" para="1" sent="2">aan</wf>
    <wf id="w9" length="7" offset="39" para="1" sent="2">termijn</wf>
    <wf id="w10" length="8" offset="47" para="1" sent="2">gebonden</wf>
    <wf id="w11" length="3" offset="57" para="2" sent="2">Net</wf>
    <wf id="w12" length="4" offset="61" para="2" sent="2">voor</wf>
    <wf id="w13" length="3" offset="66" para="2" sent="2">het</wf>
    <wf id="w14" length="8" offset="70" para="2" sent="2">verlopen</wf>
    <wf id="w15" length="3" offset="79" para="2" sent="2">van</wf>
    <wf id="w16" length="2" offset="83" para="2" sent="2">de</wf>
    <wf id="w17" length="16" offset="86" para="2" sent="2">fabrieksgarantie</wf>
    <wf id="w18" length="4" offset="103" para="2" sent="2">ging</wf>
    <wf id="w19" length="2" offset="108" para="2" sent="2">de</wf>
    <wf id="w20" length="4" offset="111" para="2" sent="2">accu</wf>
    <wf id="w21" length="3" offset="116" para="2" sent="2">van</wf>
    <wf id="w22" length="4" offset="120" para="2" sent="2">mijn</wf>
    <wf id="w23" length="5" offset="125" para="2" sent="2">Honda</wf>
    <wf id="w24" length="4" offset="131" para="2" sent="2">Jazz</wf>
    <wf id="w25" length="5" offset="136" para="2" sent="2">kapot</wf>
```

Semantic Annotation

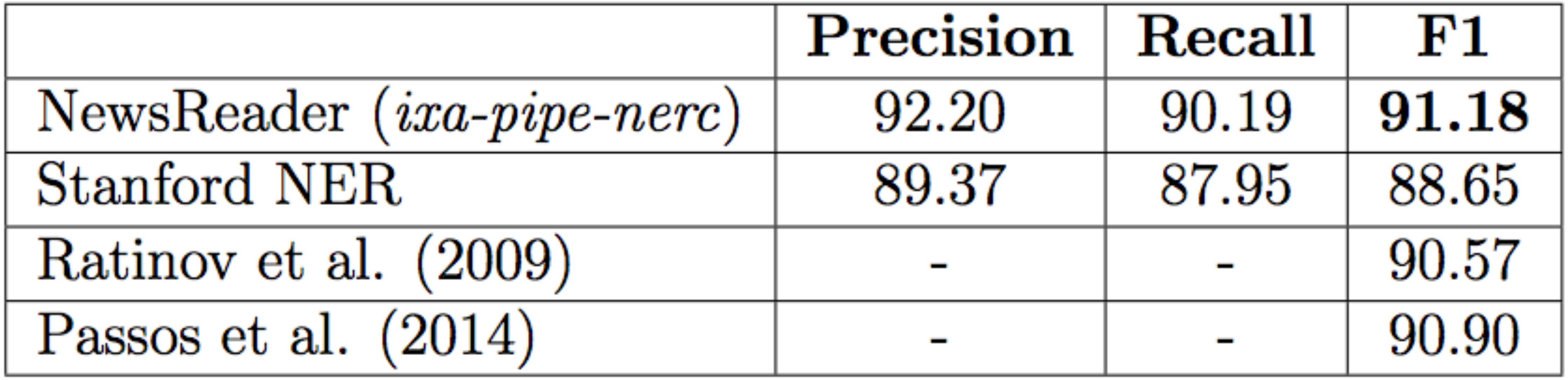
- Named Entity Recognition & Linking
- From words to concepts
- Semantic Role Labelling
- Recognising Temporal Expressions & Relations
- Wikification



Named Entity Recognition & Linking

- Semi-supervised NER: R. Agerri, G. Rigau, Robust multilingual Named Entity Recognition with shallow semi-supervised features. Artificial Intelligence, 238 (2016) 63-82. JCR 2015: 3.371

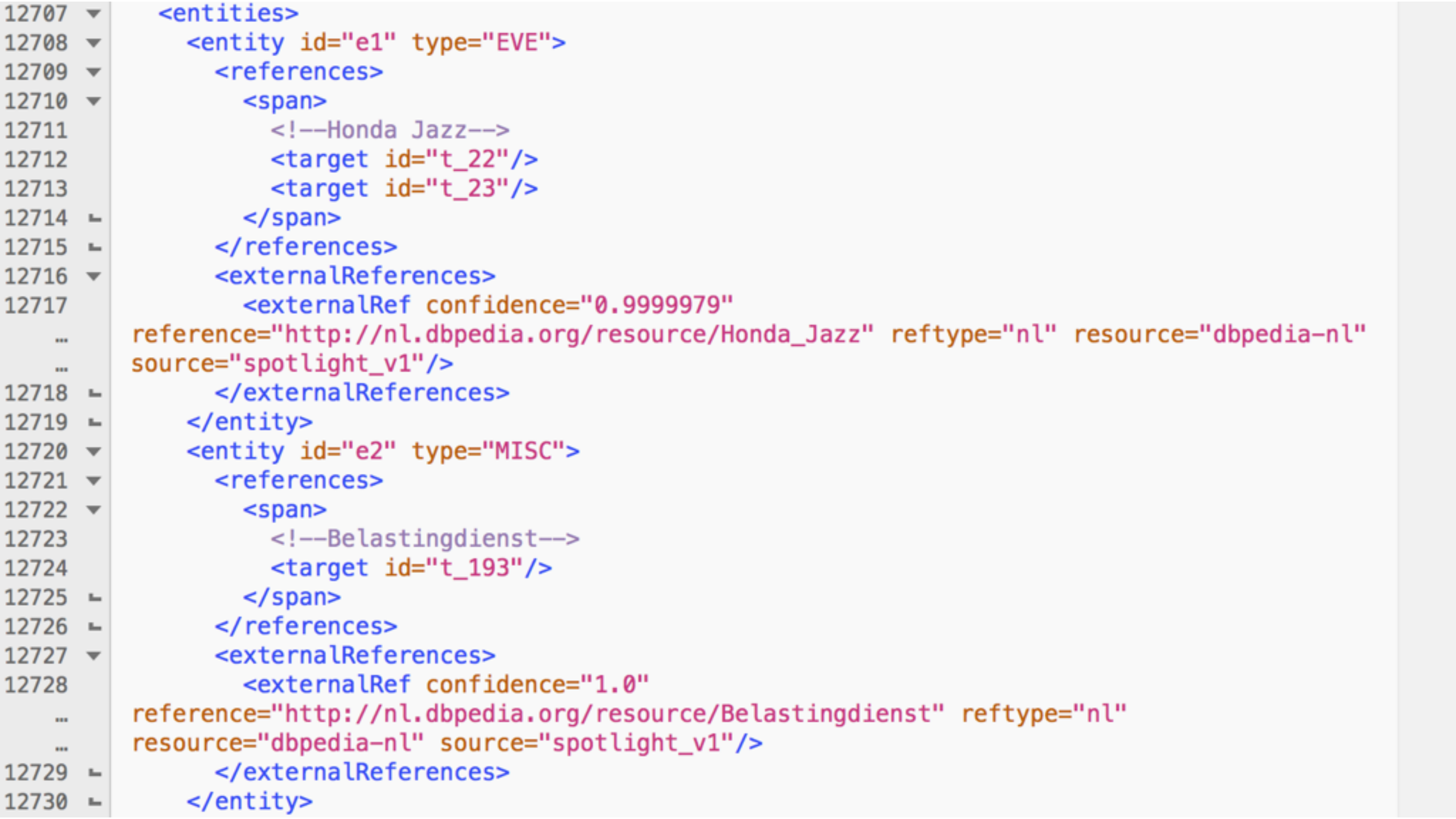
- Named Entity Linking (DBpedia Spotlight): Daiber, Joachim, et al. "Improving efficiency and accuracy in multilingual entity extraction." Proceedings of the 9th International Conference on Semantic Systems. ACM, 2013.



	Precision	**Recall**	**F1**
NewsReader (*ixa-pipe-nerc*)	92.20	90.19	**91.18**
Stanford NER	89.37	87.95	88.65
Ratinov et al. (2009)	-	-	90.57
Passos et al. (2014)	-	-	90.90

NERC CoNLL 2003 testb results.

Named Entities in NAF



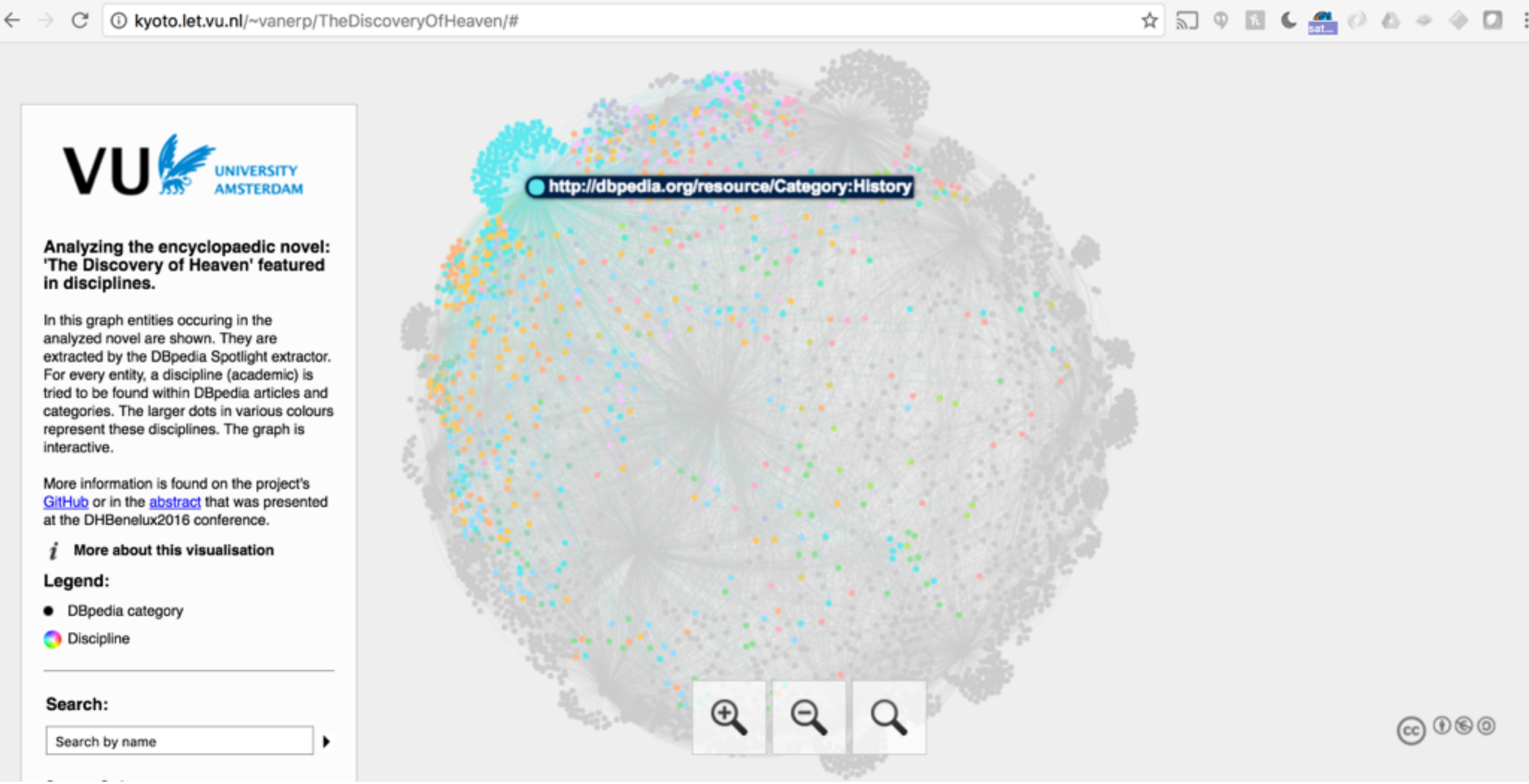
```
  <entities>
    <entity id="e1" type="EVE">
      <references>
        <span>
          <!--Honda Jazz-->
          <target id="t_22"/>
          <target id="t_23"/>
        </span>
      </references>
      <externalReferences>
        <externalRef confidence="0.9999979"
reference="http://nl.dbpedia.org/resource/Honda_Jazz" reftype="nl" resource="dbpedia-nl"
source="spotlight_v1"/>
      </externalReferences>
    </entity>
    <entity id="e2" type="MISC">
      <references>
        <span>
          <!--Belastingdienst-->
          <target id="t_193"/>
        </span>
      </references>
      <externalReferences>
        <externalRef confidence="1.0"
reference="http://nl.dbpedia.org/resource/Belastingdienst" reftype="nl"
resource="dbpedia-nl" source="spotlight_v1"/>
      </externalReferences>
    </entity>
```

Why link to a resource such as DBpedia?

- It allows you to query for fine-grained entity types: give me all politicians in the dataset, give me all football players

- Plus: the background knowledge provides additional filters: give me all politicians born after 1900 in the dataset

- Caveat: the background knowledge is not complete


kyoto.let.vu.nl/~vanerp/TheDiscoveryOfHeaven/#
http://dbpedia.org/resource/Category:History

VU UNIVERSITY AMSTERDAM

Analyzing the encyclopaedic novel: 'The Discovery of Heaven' featured in disciplines.

In this graph entities occuring in the analyzed novel are shown. They are extracted by the DBpedia Spotlight extractor. For every entity, a discipline (academic) is tried to be found within DBpedia articles and categories. The larger dots in various colours represent these disciplines. The graph is interactive.

More information is found on the project's GitHub or in the abstract that was presented at the DHBenelux2016 conference.

i **More about this visualisation**

Legend:

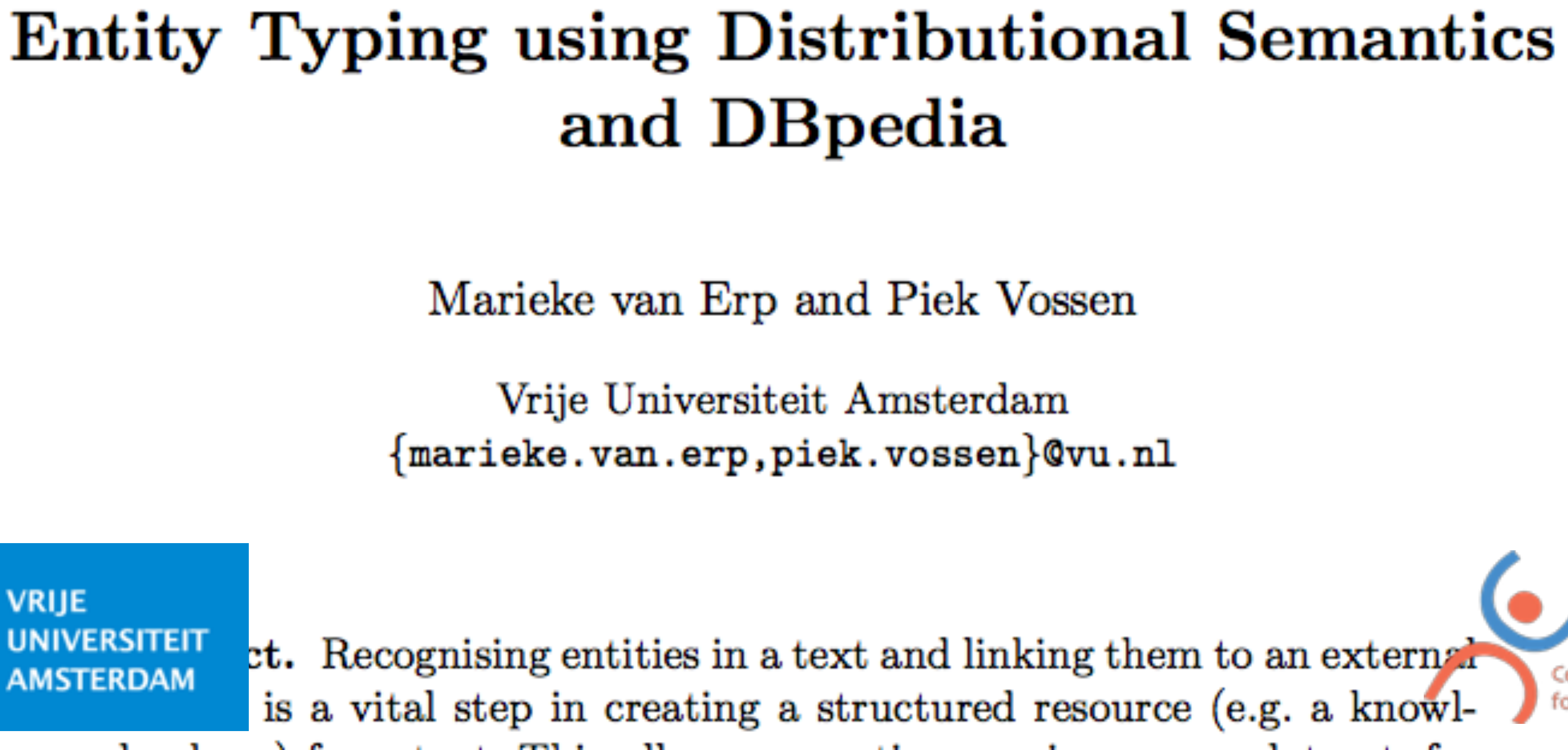
- ● DBpedia category
- ○ Discipline

Search:

Search by name ▸

Named Entity Recognition & Linking

- We are developing a new entity linker that allows for use of datasets other than DBpedia and is less sensitive to general entity popularity
- Discovering more about Dark and NIL entities is also ongoing work



Entity Typing using Distributional Semantics and DBpedia

Marieke van Erp and Piek Vossen

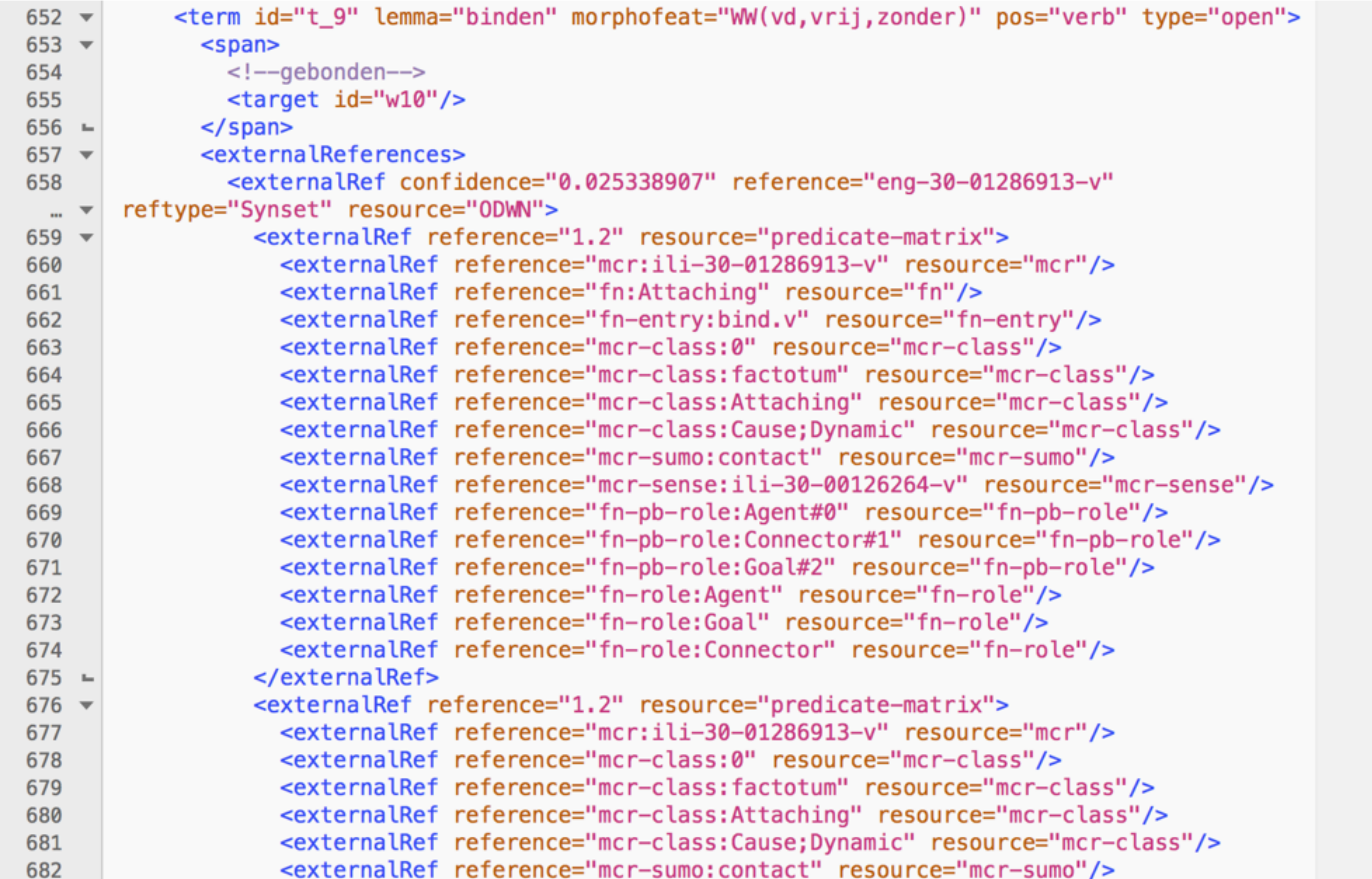
Vrije Universiteit Amsterdam
{marieke.van.erp,piek.vossen}@vu.nl

ct. Recognising entities in a text and linking them to an external is a vital step in creating a structured resource (e.g. a knowl-

From words to concepts

- Linking terms to synonyms to obtain a higher level of abstraction
- Word-sense disambiguation + WordNet + Multilingual Central Repository + Framenet + PropBank
- Stop, quit, leave, relinquish, bow out -> all linked to the concept wn:leave_office

From Words to Concepts



```
    <term id="t_9" lemma="binden" morphofeat="WW(vd,vrij,zonder)" pos="verb" type="open">
      <span>
        <!--gebonden-->
        <target id="w10"/>
      </span>
      <externalReferences>
        <externalRef confidence="0.025338907" reference="eng-30-01286913-v"
reftype="Synset" resource="ODWN">
          <externalRef reference="1.2" resource="predicate-matrix">
            <externalRef reference="mcr:ili-30-01286913-v" resource="mcr"/>
            <externalRef reference="fn:Attaching" resource="fn"/>
            <externalRef reference="fn-entry:bind.v" resource="fn-entry"/>
            <externalRef reference="mcr-class:0" resource="mcr-class"/>
            <externalRef reference="mcr-class:factotum" resource="mcr-class"/>
            <externalRef reference="mcr-class:Attaching" resource="mcr-class"/>
            <externalRef reference="mcr-class:Cause;Dynamic" resource="mcr-class"/>
            <externalRef reference="mcr-sumo:contact" resource="mcr-sumo"/>
            <externalRef reference="mcr-sense:ili-30-00126264-v" resource="mcr-sense"/>
            <externalRef reference="fn-pb-role:Agent#0" resource="fn-pb-role"/>
            <externalRef reference="fn-pb-role:Connector#1" resource="fn-pb-role"/>
            <externalRef reference="fn-pb-role:Goal#2" resource="fn-pb-role"/>
            <externalRef reference="fn-role:Agent" resource="fn-role"/>
            <externalRef reference="fn-role:Goal" resource="fn-role"/>
            <externalRef reference="fn-role:Connector" resource="fn-role"/>
          </externalRef>
          <externalRef reference="1.2" resource="predicate-matrix">
            <externalRef reference="mcr:ili-30-01286913-v" resource="mcr"/>
            <externalRef reference="mcr-class:0" resource="mcr-class"/>
            <externalRef reference="mcr-class:factotum" resource="mcr-class"/>
            <externalRef reference="mcr-class:Attaching" resource="mcr-class"/>
            <externalRef reference="mcr-class:Cause;Dynamic" resource="mcr-class"/>
            <externalRef reference="mcr-sumo:contact" resource="mcr-sumo"/>
```

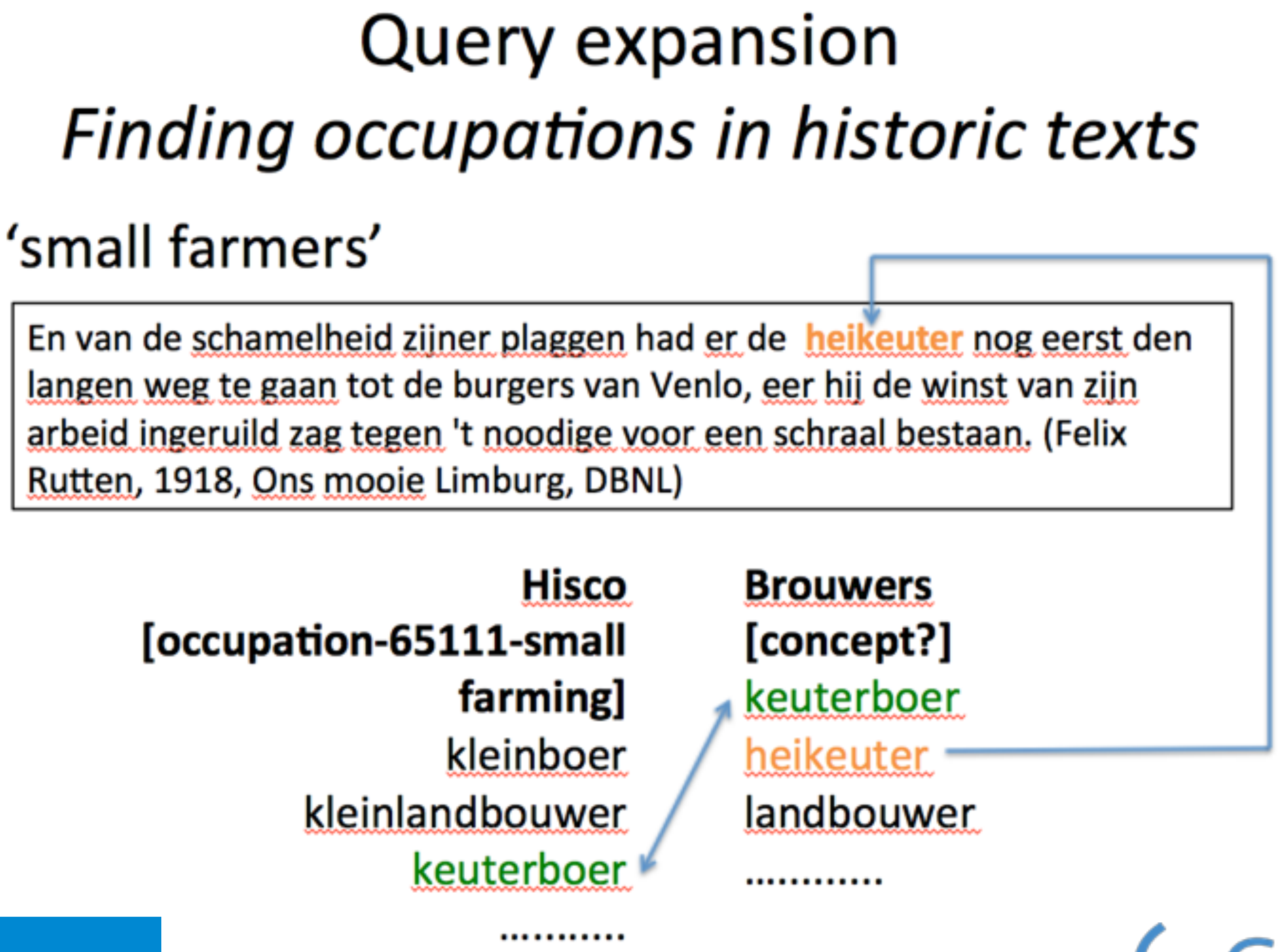
Why link to WordNet/ConceptNet/etc?

- It allows you to query for types rather than instances: give me all lawsuits in the dataset
- In the context of CLARIAH, we are converting various diachronous lexicons to Linked Data
  - integrate resources
  - tag interesting concepts in text
  - query expansion

VRIJE UNIVERSITEIT AMSTERDAM

CLARIAH Common Lab Research Infrastructure for the Arts and Humanities

New synonym/concept lists are easy to plug in



Query expansion

Finding occupations in historic texts

‘small farmers’

En van de schamelheid zijner plaggen had er de heikeuter nog eerst den langen weg te gaan tot de burgers van Venlo, eer hij de winst van zijn arbeid ingeruild zag tegen 't noodige voor een schraal bestaan. (Felix Rutten, 1918, Ons mooie Limburg, DBNL)

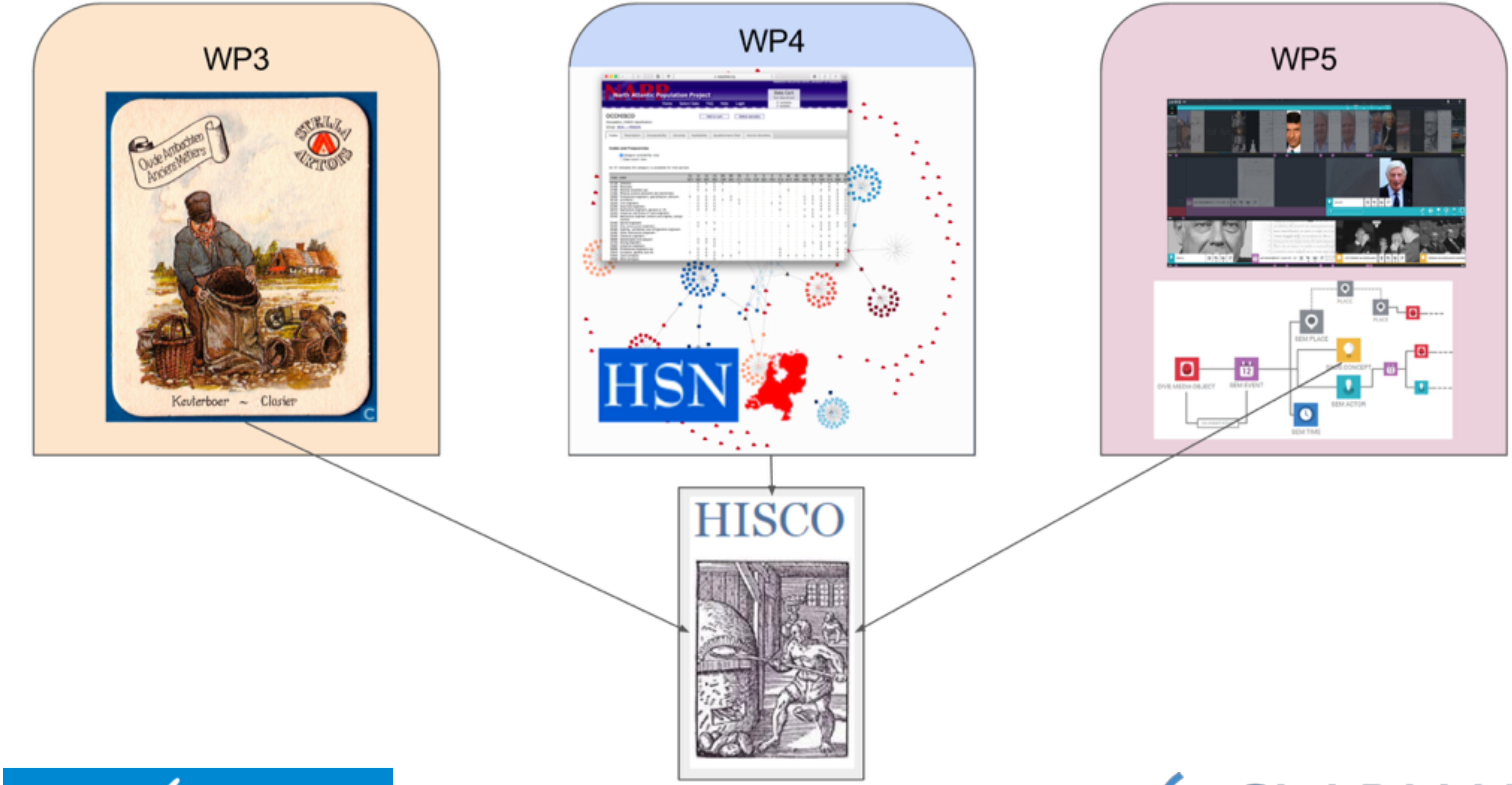
Hisco
[occupation-65111-small farming]
kleinboer
kleinlandbouwer
keuterboer
....

Brouwers
[concept?]
keuterboer
heikeuter
landbouwer
....

VRIJE UNIVERSITEIT AMSTERDAM

CLARIAH
Common Lab Research Infrastructure for the Arts and Humanities

New synonym/concept lists are easy to plug in

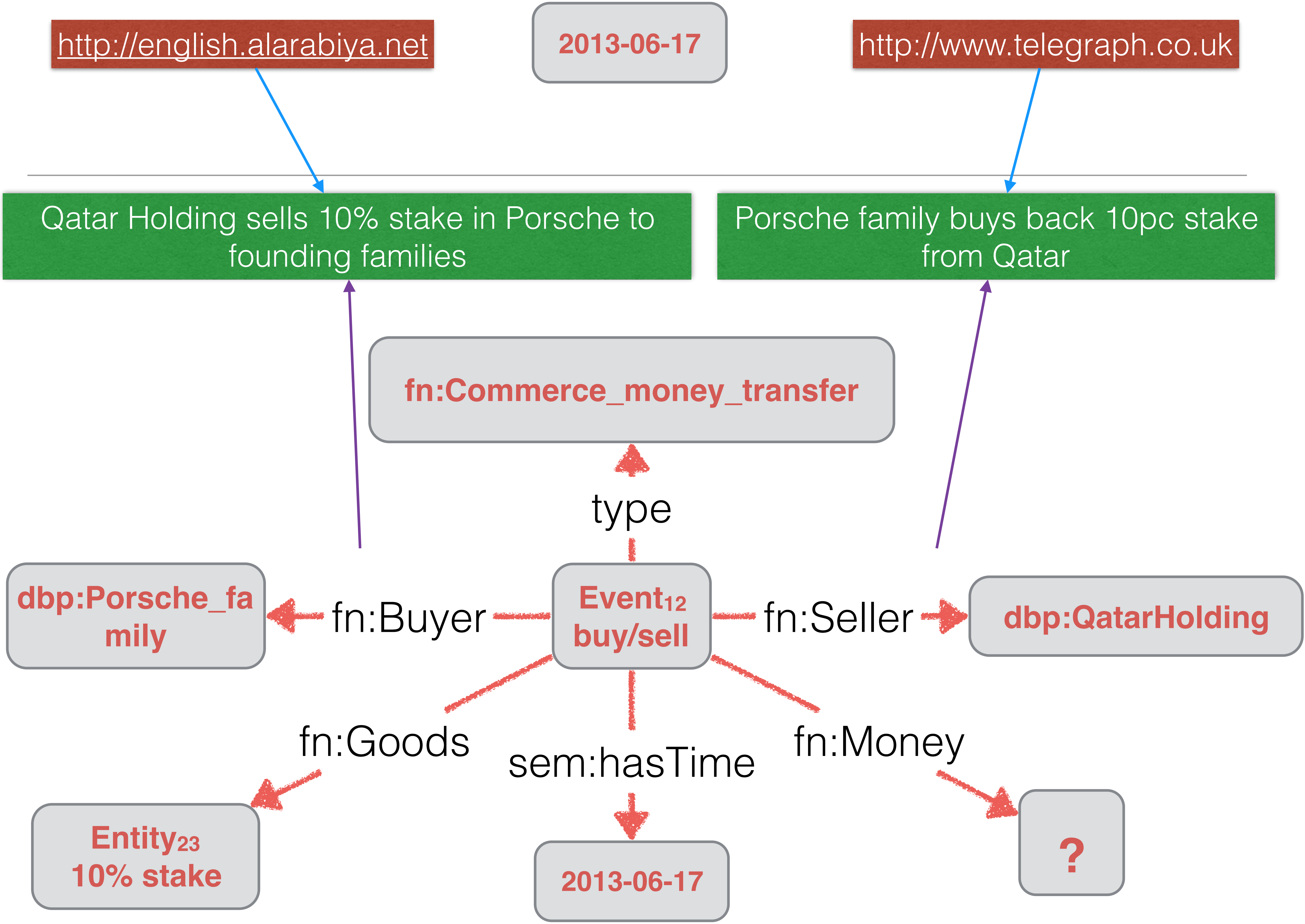


VRIJE UNIVERSITEIT AMSTERDAM

CLARIAH
Common Lab Research Infrastructure for the Arts and Humanities

Semantic Role Labelling

- Detecting the agent, patient, recipient and theme of a sentence
  - Mary sold the book to John
  - Agent: Mary
  - Recipient: John
  - Theme: the book



http://english.alarabiya.net

2013-06-17

http://www.telegraph.co.uk

Qatar Holding sells 10% stake in Porsche to founding families

Porsche family buys back 10pc stake from Qatar

fn:Commerce_money_transfer

type

dbp:Porsche_family

fn:Buyer

Event$_{12}$ buy/sell

fn:Seller

dbp:QatarHolding

fn:Goods

sem:hasTime

fn:Money

Entity$_{23}$ 10% stake

2013-06-17

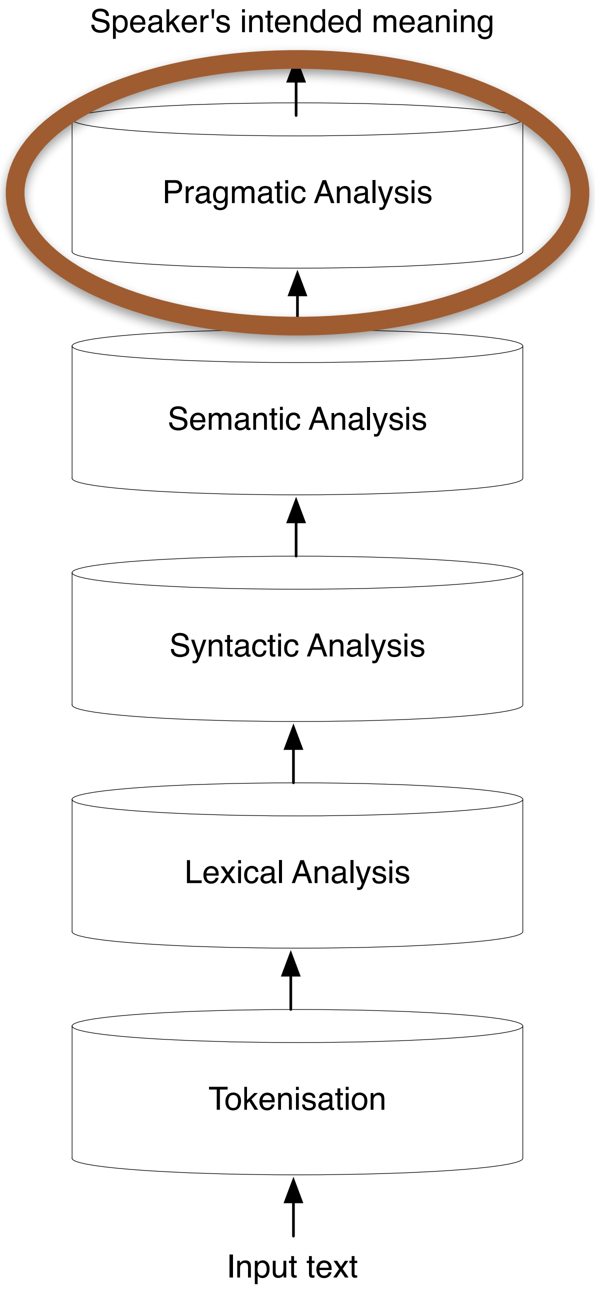
?

Event abstractions

- Enable searches such as: Give me all lawsuits in which a politician was involved between 1990 and 2000.

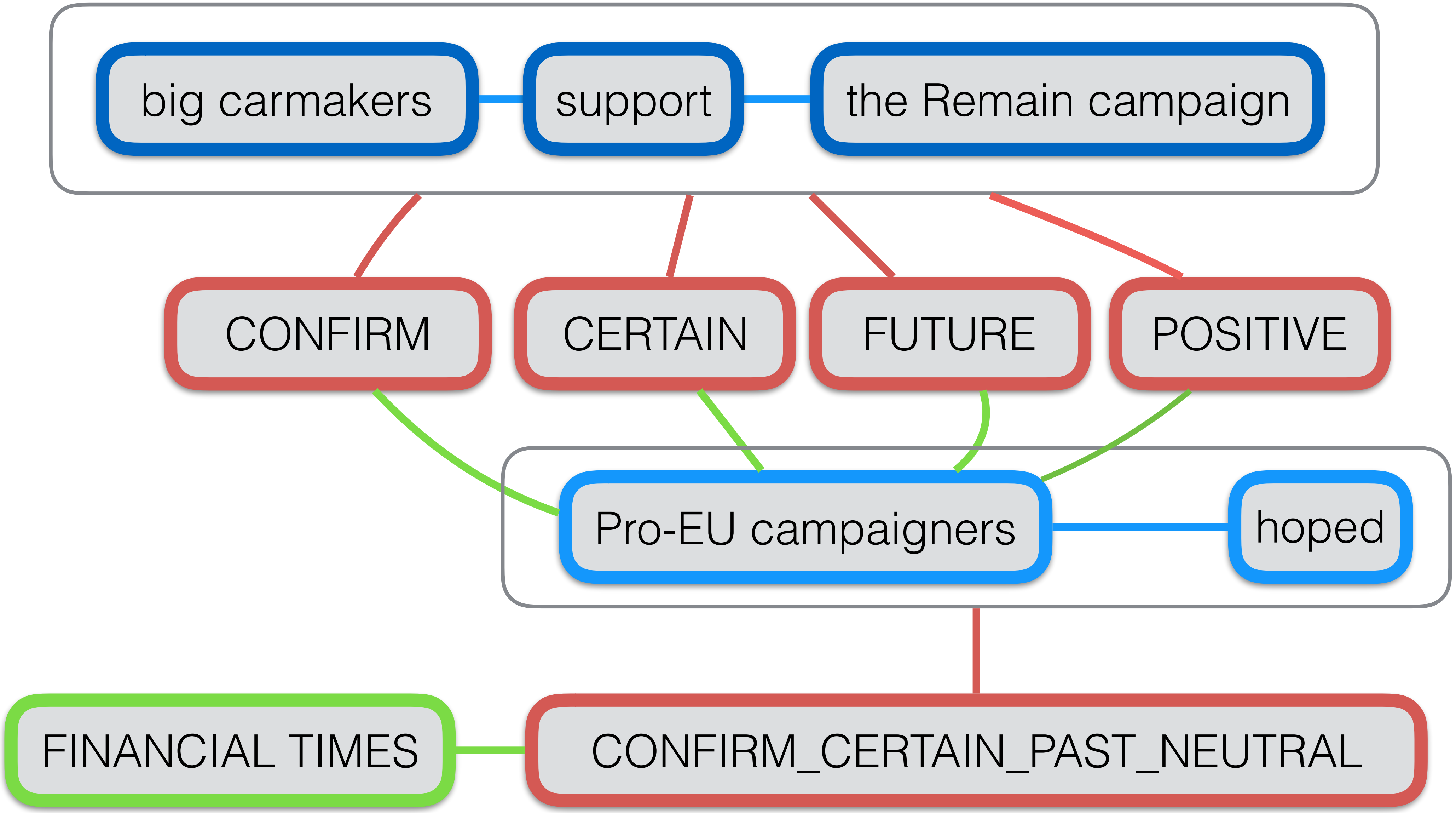
Pragmatic Analysis

- Factuality/Attribution
    - Who said what, who agrees with whom, how certain is a speaker about her statement, is she talking about the past, present or future?



Perspective

Pro-EU campaigners have hoped that big carmakers would also support the Remain campaign.



big carmakers

support

the Remain campaign

CONFIRM

CERTAIN

FUTURE

POSITIVE

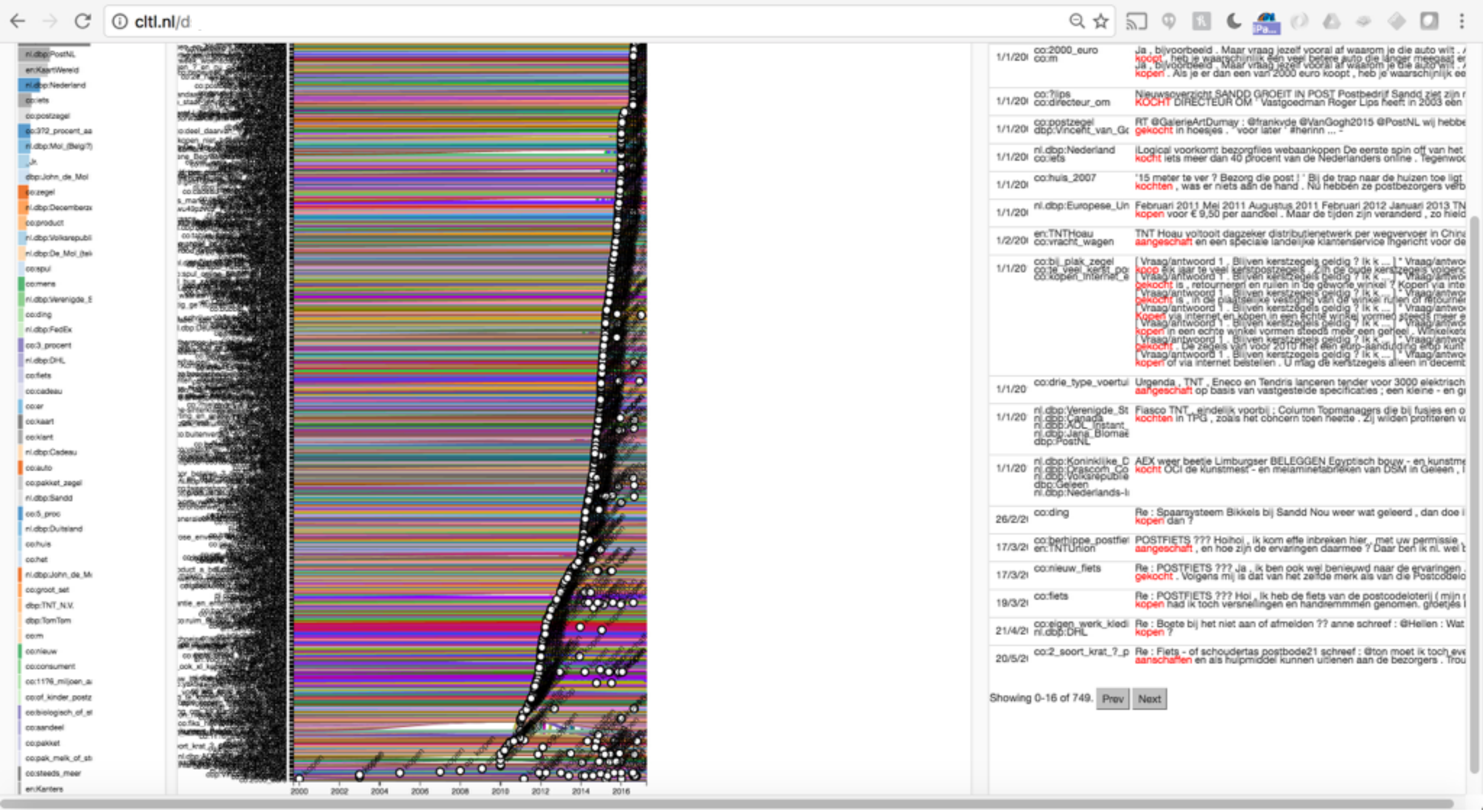
Pro-EU campaigners

hoped

FINANCIAL TIMES

CONFIRM_CERTAIN_PAST_NEUTRAL

and beyond…



1/1/20	co:2000_euro co:m	Ja , bijvoorbeeld . Maar vraag jezelf vooral af waarom je die auto wilt . / koopt , heb je waarschijnlijk een veel betere auto die langer meegaat er Ja , bijvoorbeeld . Maar vraag jezelf vooral af waarom je die auto wilt . / kopen . Als je er dan een van 2000 euro koopt , heb je waarschijnlijk ee
1/1/20	co:?lips co:directeur_om	Nieuwsoverzicht SANDD GROEIT IN POST Postbedrijf Sandd ziet zijn r KOCHT DIRECTEUR OM ' Vastgoedman Roger Lips heeft in 2003 een
1/1/20	co:postzegel dbp:Vincent_van_Gc	RT @GalerieArtDumay : @frankvde @VanGogh2015 @PostNL wij hebbe gekocht in hoesjes . ' voor later ' #herinn ... -
1/1/20	nl.dbp:Nederland co:iets	iLogical voorkomt bezorgfiles webaankopen De eerste spin off van het kocht iets meer dan 40 procent van de Nederlanders online . Tegenwoo
1/1/20	co:huis_2007	'15 meter te ver ? Bezorg die post ! ' Bij de trap naar de huizen toe ligt kochten , was er niets aan de hand . Nu hebben ze postbezorgers verb
1/1/20	nl.dbp:Europese_Un	Februari 2011 Mei 2011 Augustus 2011 Februari 2012 Januari 2013 TN kopen voor € 9,50 per aandeel . Maar de tijden zijn veranderd , zo hield
1/2/20	en:TNTHoau co:vracht_wagen	TNT Hoau voltooit dagzeker distributienetwerk per wegvervoer in China aangeschaft en een speciale landelijke klantenservice ingericht voor de
1/1/20	co:bij_plak_zegel co:te_veel_kerst_po co:kopen_internet_e	[Vraag/antwoord 1 . Blijven kerstzegels geldig ? Ik k ...] " Vraag/antwo koop elk jaar te veel kerstpostzegels . Zijn de oude kerstzegels volgend [Vraag/antwoord 1 . Blijven kerstzegels geldig ? Ik k ...] " Vraag/antwo gekocht is , retourneren en ruilen in de gewone winkel ? Kopen via inte [Vraag/antwoord 1 . Blijven kerstzegels geldig ? Ik k ...] " Vraag/antwo gekocht is , in de plaatselijke vestiging van de winkel ruilen of retourner [Vraag/antwoord 1 . Blijven kerstzegels geldig ? Ik k ...] " Vraag/antwo Kopen via internet en kopen in een echte winkel vormen steeds meer e [Vraag/antwoord 1 . Blijven kerstzegels geldig ? Ik k ...] " Vraag/antwo kopen in een echte winkel vormen steeds meer een geheel . Winkelkete [Vraag/antwoord 1 . Blijven kerstzegels geldig ? Ik k ...] " Vraag/antwo gekocht . De zegels van voor 2010 met een euro-aanduiding erop kunt [Vraag/antwoord 1 . Blijven kerstzegels geldig ? Ik k ...] " Vraag/antwo kopen of via internet bestellen . U mag de kerstzegels alleen in decemt
1/1/20	co:drie_type_voertui	Urgenda , TNT , Eneco en Tendris lanceren tender voor 3000 elektrisch aangeschaft op basis van vastgestelde specificaties ; een kleine - en gr
1/1/20	nl.dbp:Verenigde_St nl.dbp:Canada nl.dbp:AOL_Instant nl.dbp:Jana_Blomae dbp:PostNL	Fiasco TNT , eindelijk voorbij ; Column Topmanagers die bij fusies en o kochten in TPG , zoals het concern toen heette . Zij wilden profiteren va
1/1/20	nl.dbp:Koninklijke_D nl.dbp:Orascom_Co nl.dbp:Volksrepublie dbp:Geleen nl.dbp:Nederlands-I	AEX weer beetje Limburgser BELEGGEN Egyptisch bouw - en kunstme kocht OCI de kunstmest - en melaminefabrieken van DSM in Geleen , I
26/2/2	co:ding	Re : Spaarsysteem Bikkels bij Sandd Nou weer wat geleerd , dan doe i kopen dan ?
17/3/2	co:berhippe_postfiet en:TNTUnion	POSTFIETS ??? Hoihoi , ik kom effe inbreken hier , met uw permissie , aangeschaft , en hoe zijn de ervaringen daarmee ? Daar ben ik nl. wel b
17/3/2	co:nieuw_fiets	Re : POSTFIETS ??? Ja , ik ben ook wel benieuwd naar de ervaringen . gekocht . Volgens mij is dat van het zelfde merk als van die Postcodelo
19/3/2	co:fiets	Re : POSTFIETS ??? Hoi , ik heb de fiets van de postcodeloterij (mijn r kopen had ik toch versnellingen en handremmmen genomen. groetjes I
21/4/2	co:eigen_werk_kledi nl.dbp:DHL	Re : Boete bij het niet aan of afmelden ?? anne schreef : @Hellen : Wat kopen ?
20/5/2	co:2_soort_krat_?_p	Re : Fiets - of schoudertas postbode21 schreef : @ton moet ik toch eve aanschaffen en als hulpmiddel kunnen uitlenen aan de bezorgers . Trou

Showing 0-16 of 749. Prev Next

Find out more

- All modules and evaluations are described in: http://kyoto.let.vu.nl/newsreader_deliverables/NWR-D4-2-3.pdf (158 pages!)
- http://www.newsreader-project.eu/results/software/
    - Black box setup
    - Links to individual modules on Github
    - Hadoop package for batch processing
- New developments: http://www.clariah.nl & https://github.com/clariah

VU VRIJE UNIVERSITEIT AMSTERDAM

CLARIAH
Common Lab Research Infrastructure
for the Arts and Humanities

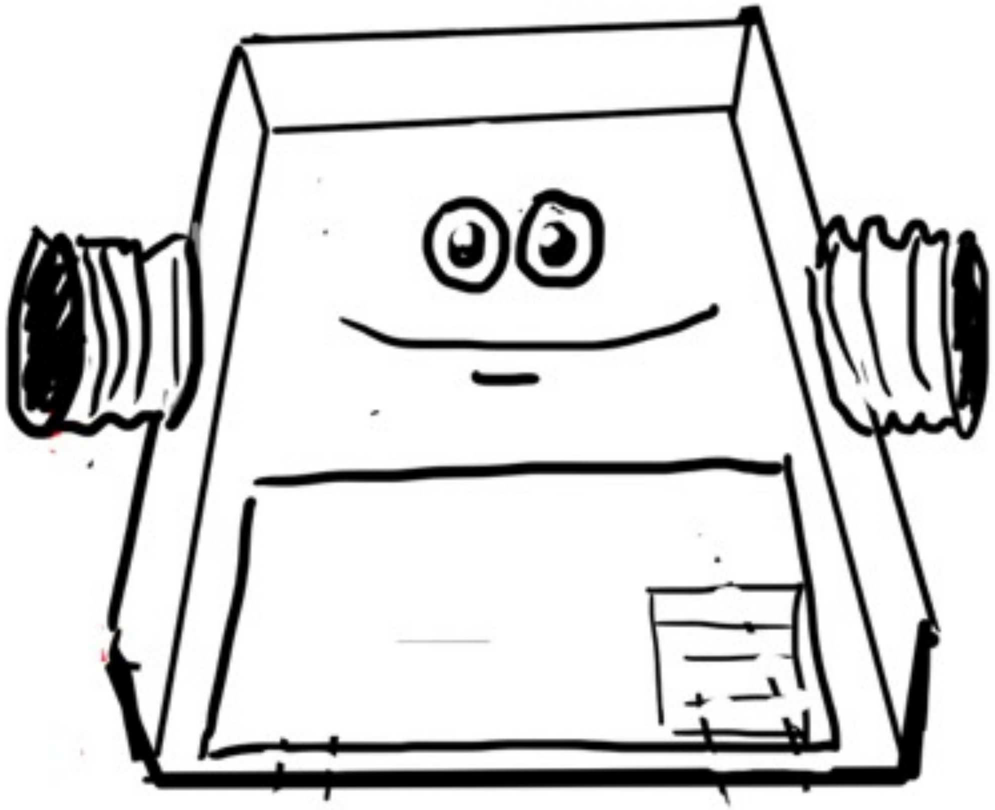
Discussion

- It’s research software (no fancy interface)
- Currently not adapted to deal with old spelling variants/OCR/etc
- NLP isn’t perfect (but humans don’t always agree either!)
- What would it take for you to start using such tools?
- What types of analyses are most interesting to the community?
- What use cases are most useful to the community at this point in time?

VRIJE UNIVERSITEIT AMSTERDAM

CLARIAH Common Lab Research Infrastructure for the Arts and Humanities

Thank you for your attention



https://youtu.be/rYLaVN3oqLI

VU VRIJE UNIVERSITEIT AMSTERDAM

CLARIAH Common Lab Research Infrastructure for the Arts and Humanities