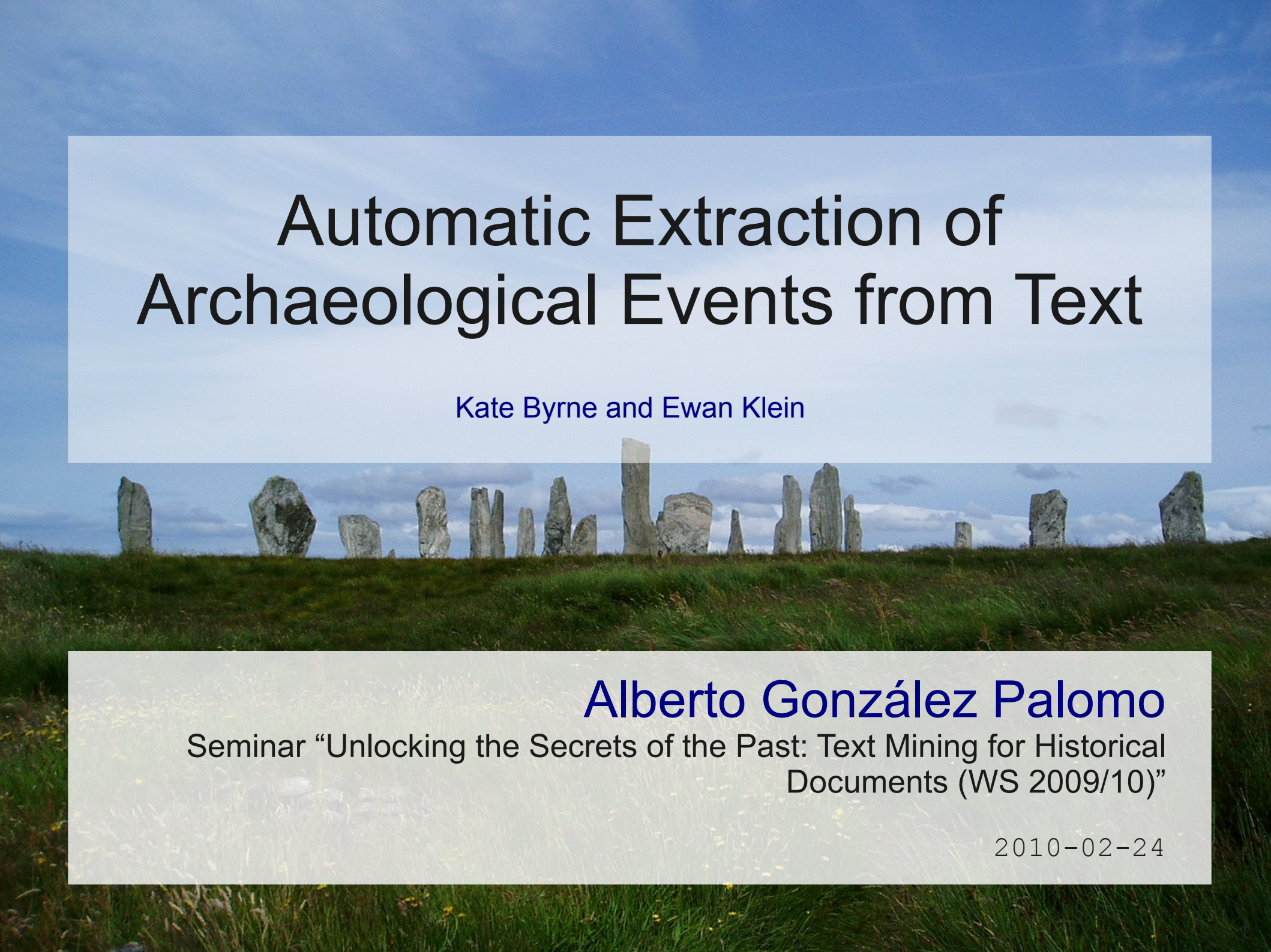


# Automatic Extraction of Archaeological Events from Text

Kate Byrne and Ewan Klein

Alberto González Palomo

Seminar “Unlocking the Secrets of the Past: Text Mining for Historical Documents (WS 2009/10)”

2010-02-24

Site456

SOUTH WALLS, MISBISTER, THE LOFTS
ND38NW29 centred 3325 8885

Sites recorded during an archaeological survey undertaken on the lands of the Loft, Longhope, as part of the pilot scheme for the Historic Scotland Farm Ancient Monument Survey Grant Scheme. 8885 Cairn. ND 3339 8886 Clearance cairn. ND 3342 8884 Sub-rectangular cairn. ND 3339 8883 Well Sponsors: Historic Scotland, M J Jones. N Card 1998

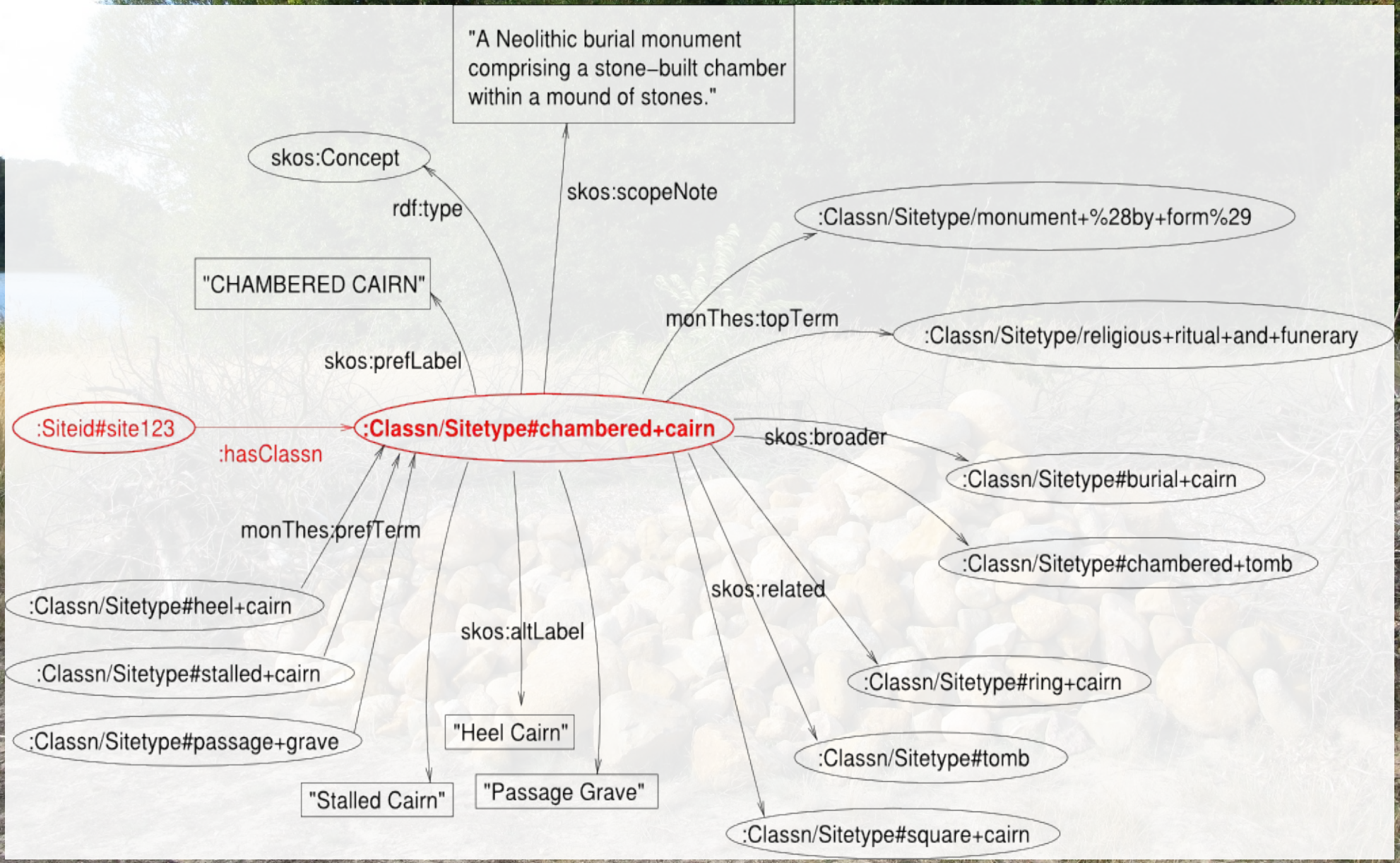
Clearance cairn in Germany

Memorial cairn in South Africa

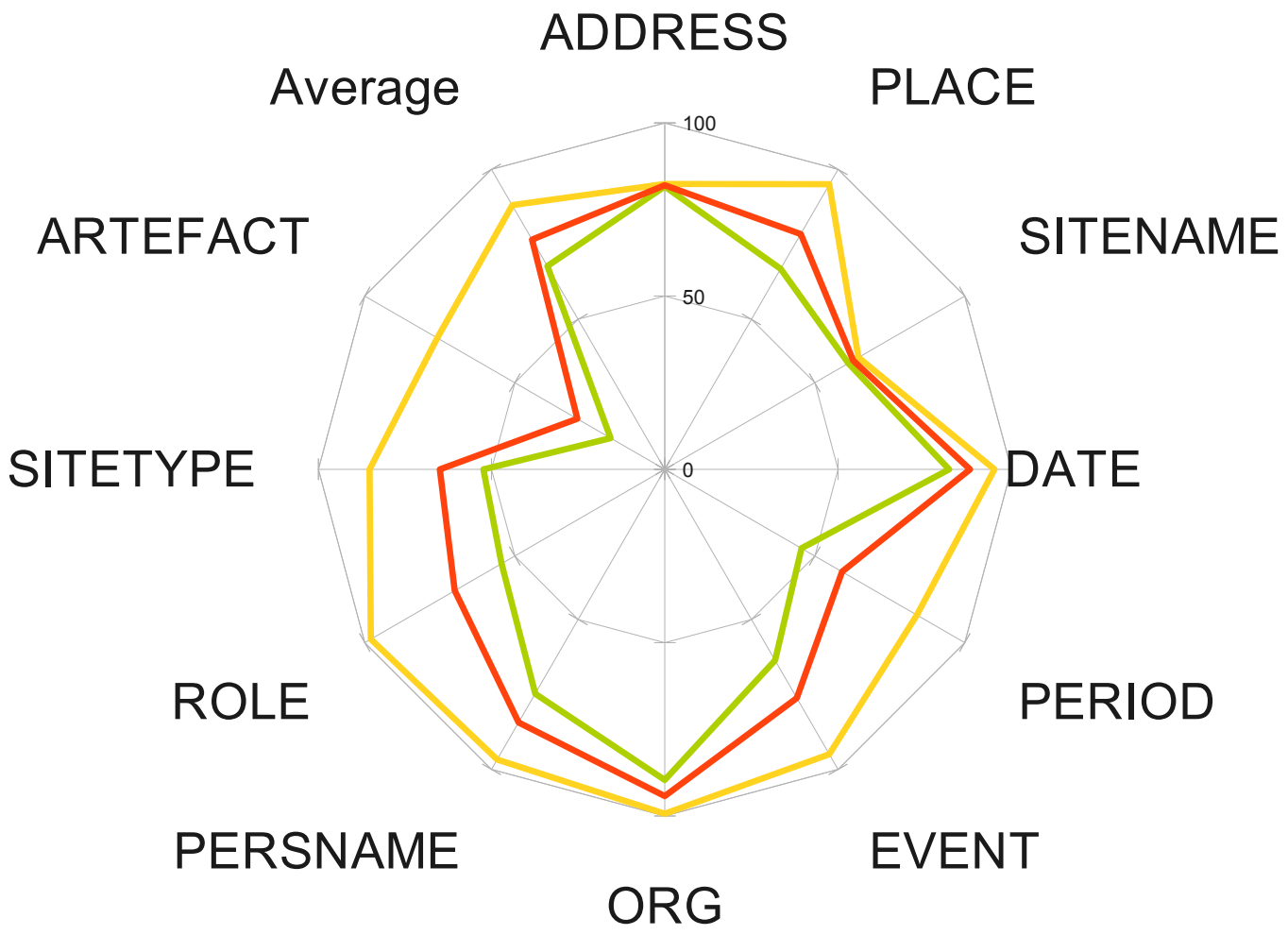
Site456

SOUTH WALLS, MISBISTER, THE LOFTS
ND38NW29 centred 3325 8885

Sites recorded during an archaeological survey undertaken on the lands of the Loft, Longhope, as part of the pilot scheme for the Historic Scotland Farm Ancient Monument Survey Grant Scheme. 8885 Cairn. ND 3339 8886 Clearance cairn. ND 3342 8884 Sub-rectangular cairn. ND 3339 8883 Well Sponsors: Historic Scotland, M J Jones. N Card 1998

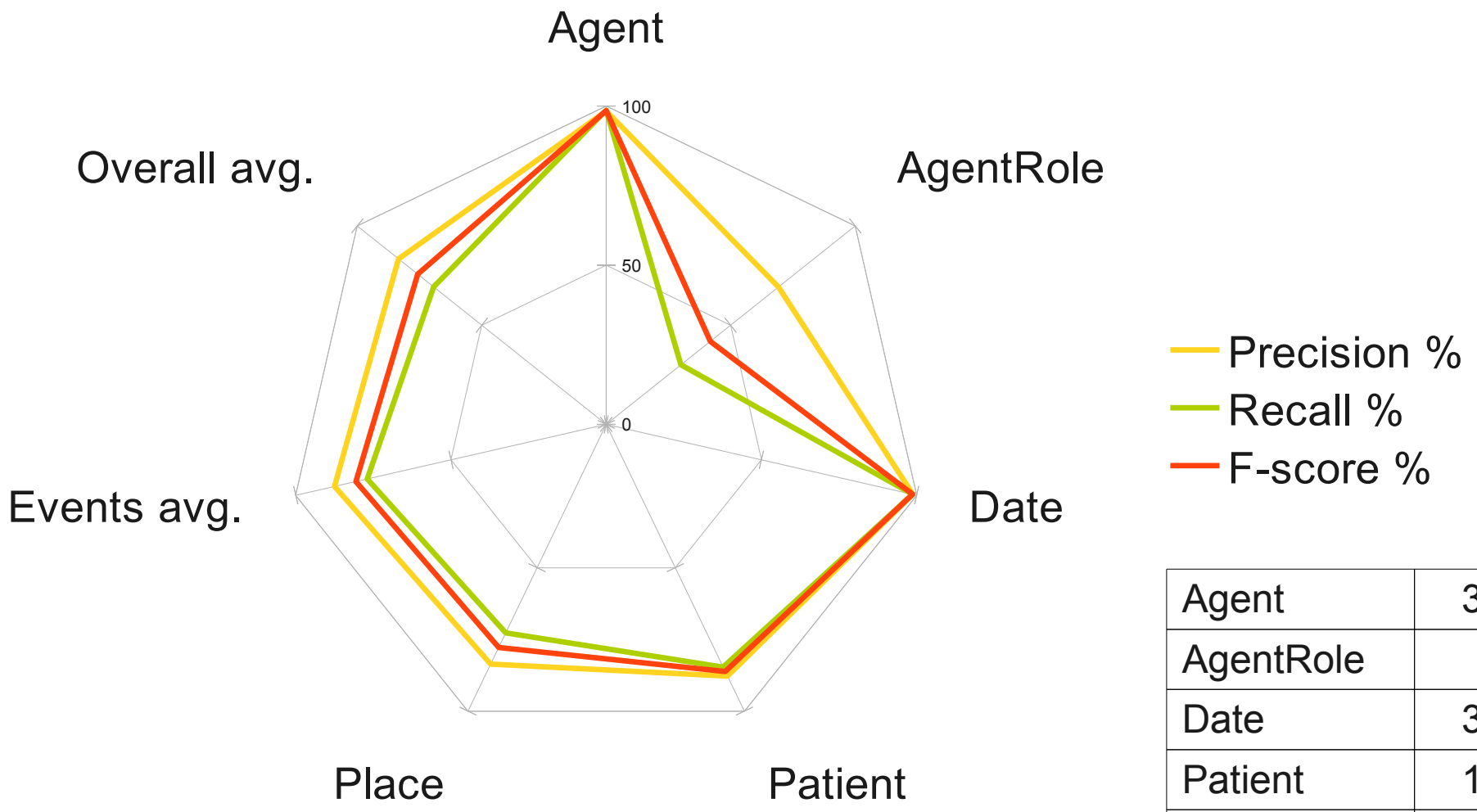

"A Neolithic burial monument comprising a stone–built chamber within a mound of stones."
skos:Concept
rdf:type
skos:scopeNote
:Classn/Sitetype/monument+%28by+form%29
"CHAMBERED CAIRN"
monThes:topTerm
:Classn/Sitetype/religious+ritual+and+funerary
skos:prefLabel
:Siteid#site123
:Classn/Sitetype#chambered+cairn
skos:broader
:hasClassn
:Classn/Sitetype#burial+cairn
monThes:prefTerm
:Classn/Sitetype#chambered+tomb
skos:related
:Classn/Sitetype#heel+cairn
skos:altLabel
:Classn/Sitetype#stalled+cairn
:Classn/Sitetype#ring+cairn
:Classn/Sitetype#passage+grave
"Heel Cairn"
:Classn/Sitetype#tomb
"Stalled Cairn"
"Passage Grave"
:Classn/Sitetype#square+cairn

# Results: Named Entity Recognition



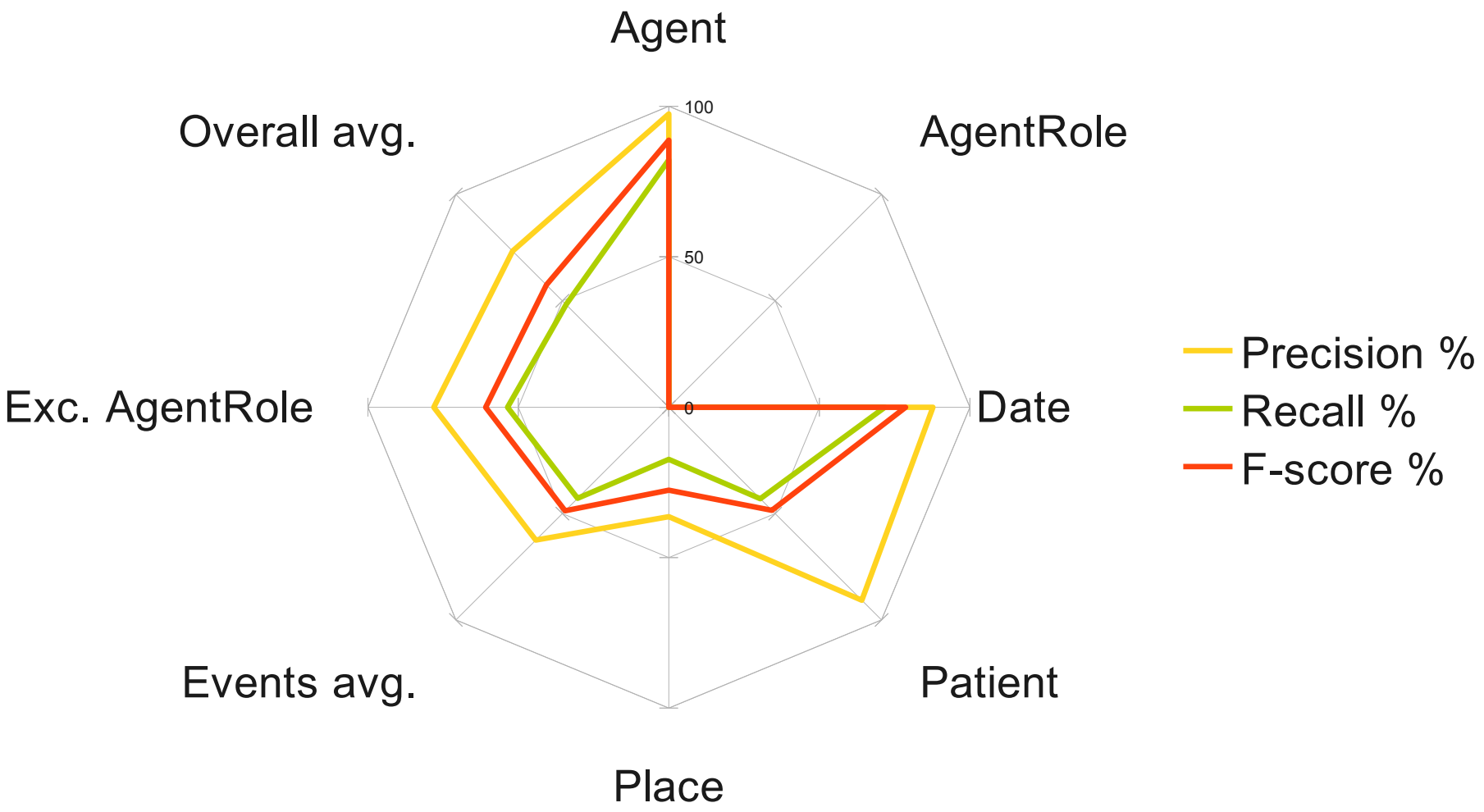

| | |
|---|---|
| ADDRESS | 3458 |
| PLACE | 2503 |
| SITENAME | 1712 |
| DATE | 3519 |
| PERIOD | 400 |
| EVENT | 3176 |
| ORG | 2730 |
| PERSNAME | 2318 |
| ROLE | 90 |
| SITETYPE | 5668 |
| ARTEFACT | 879 |
| Total | 27453 |

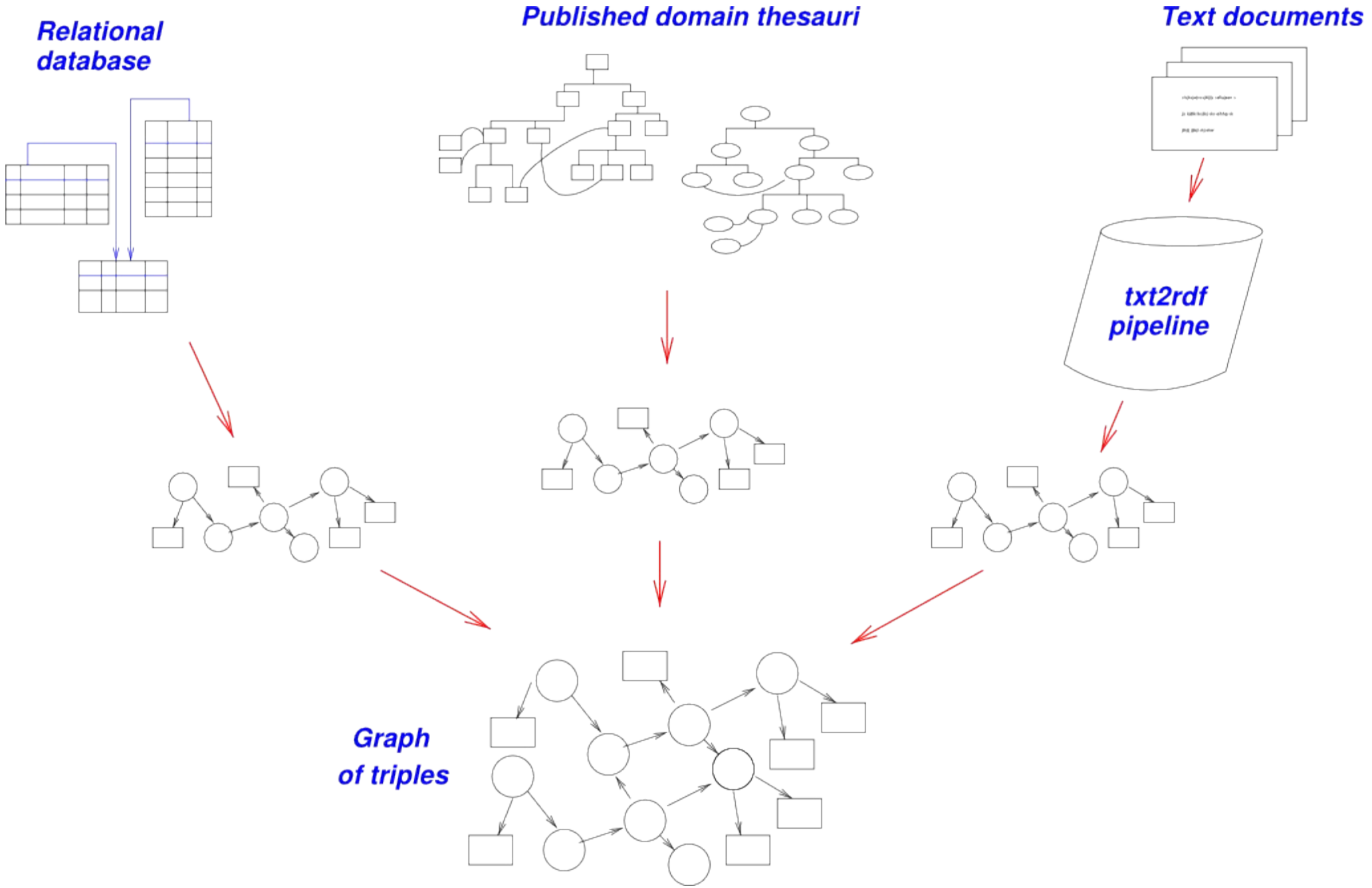
# Results: Relation Extraction



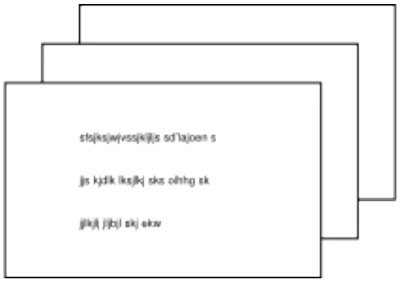

| | |
|---|---|
| Agent | 3794 |
| AgentRole | 13 |
| Date | 3189 |
| Patient | 1553 |
| Place | 341 |
| Events | 8890 |
| Overall | 21932 |

# Results: whole pipeline NER + RE

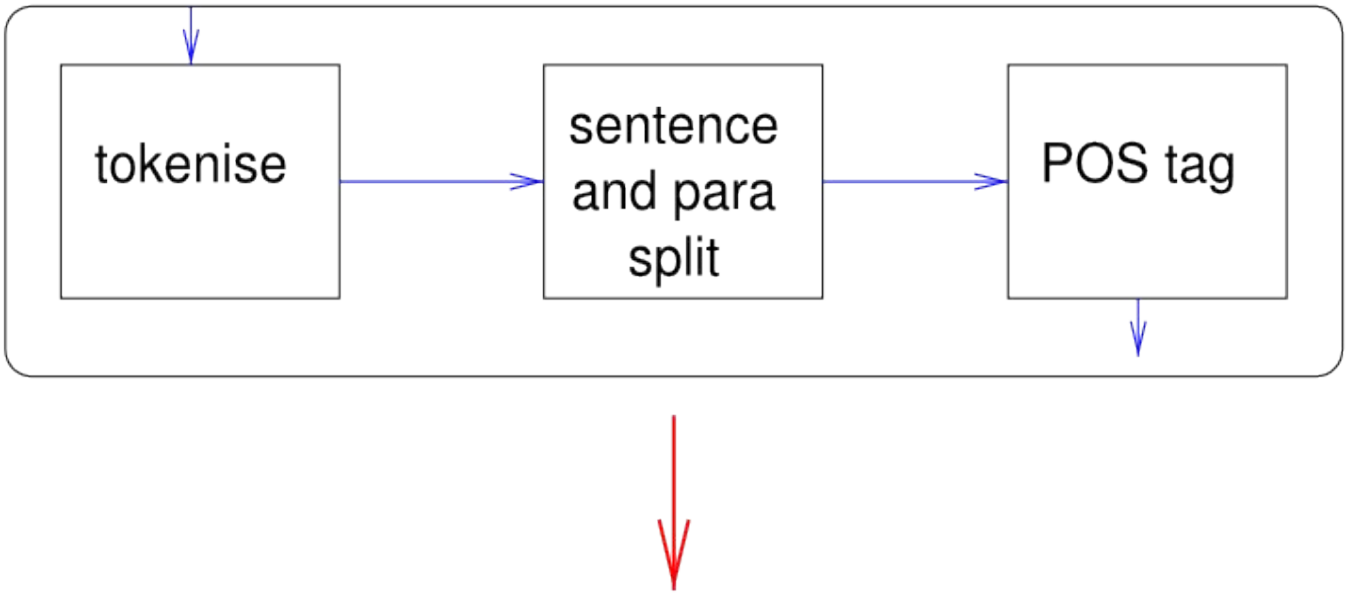


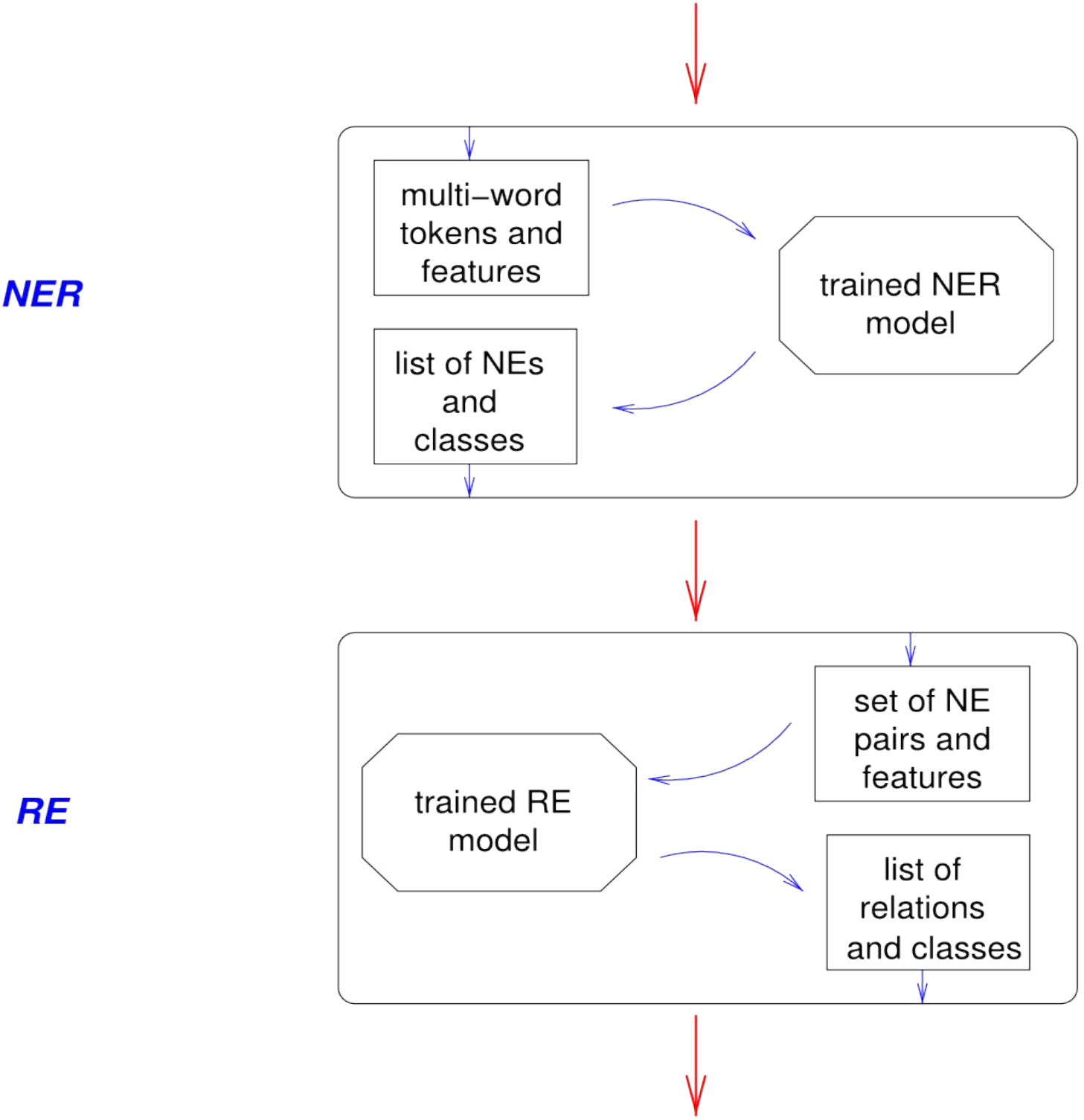

Relational database
Published domain thesauri
Text documents
txt2rdf pipeline
Graph of triples

***Text documents***

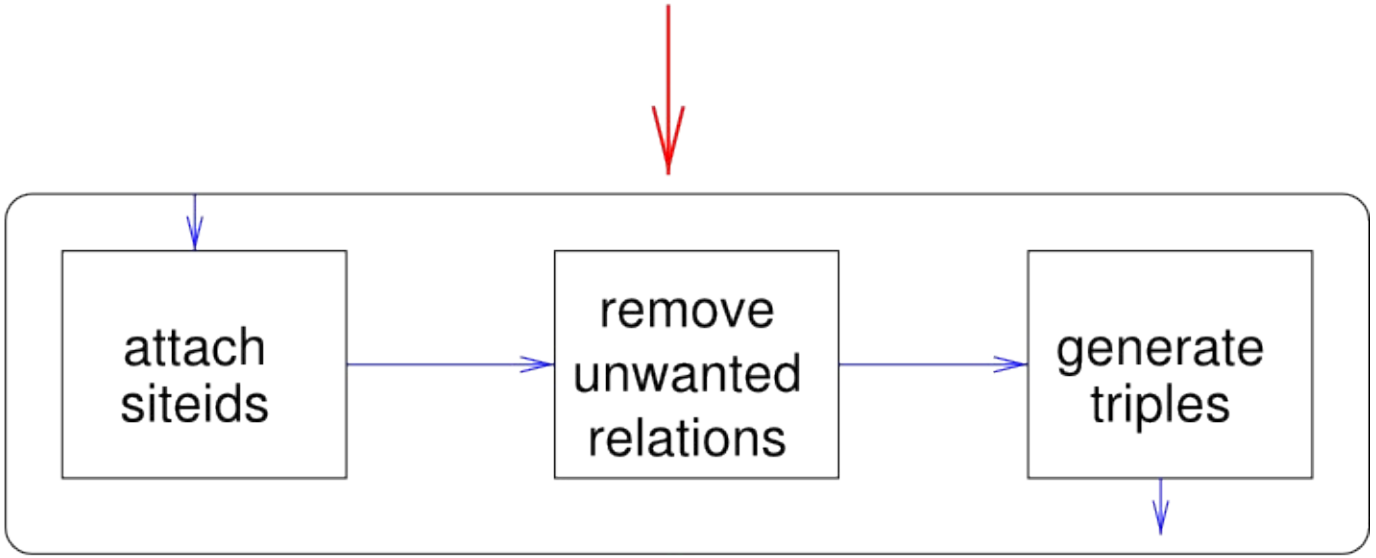


***Pre–processing***

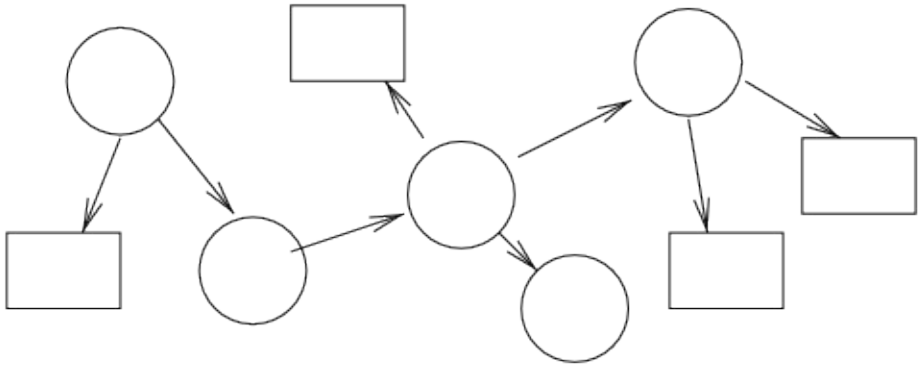



NER
multi–word tokens and features
trained NER model
list of NEs and classes
RE
set of NE pairs and features
trained RE model
list of relations and classes

***RDF translation***




***Graph of triples***



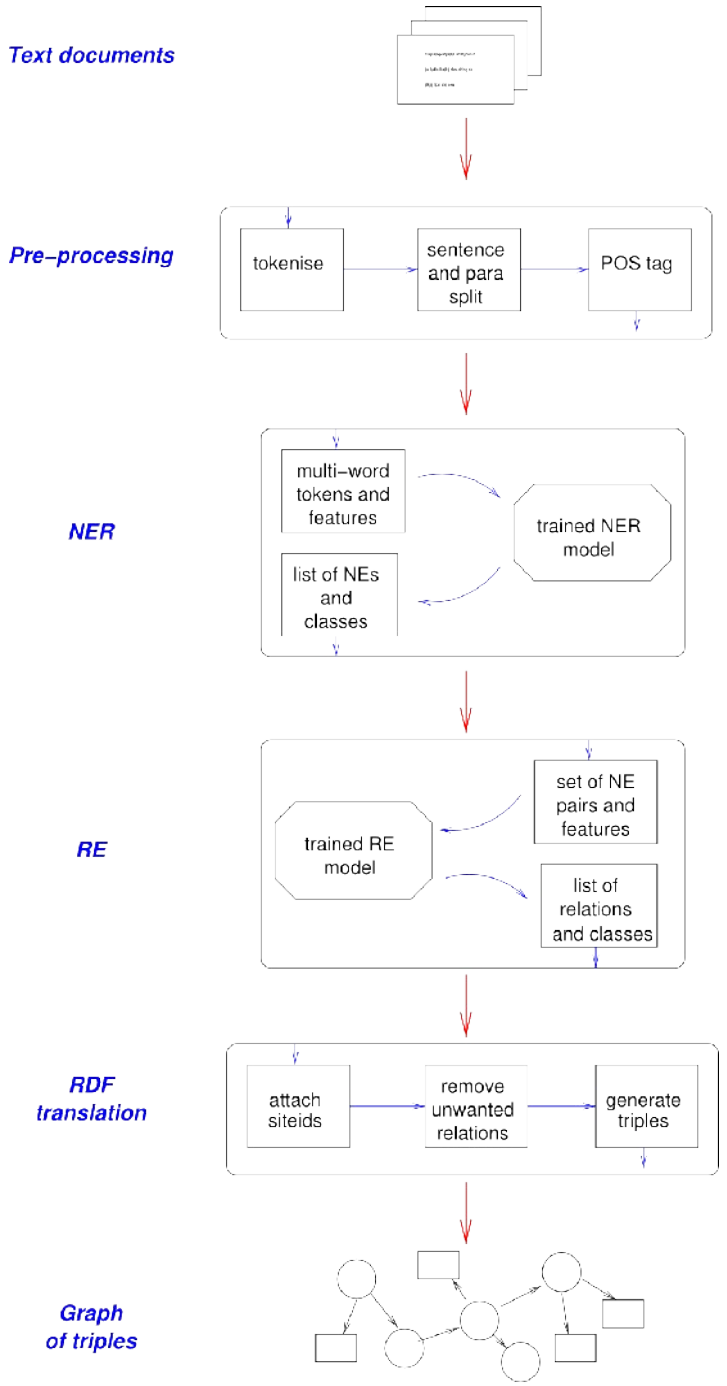
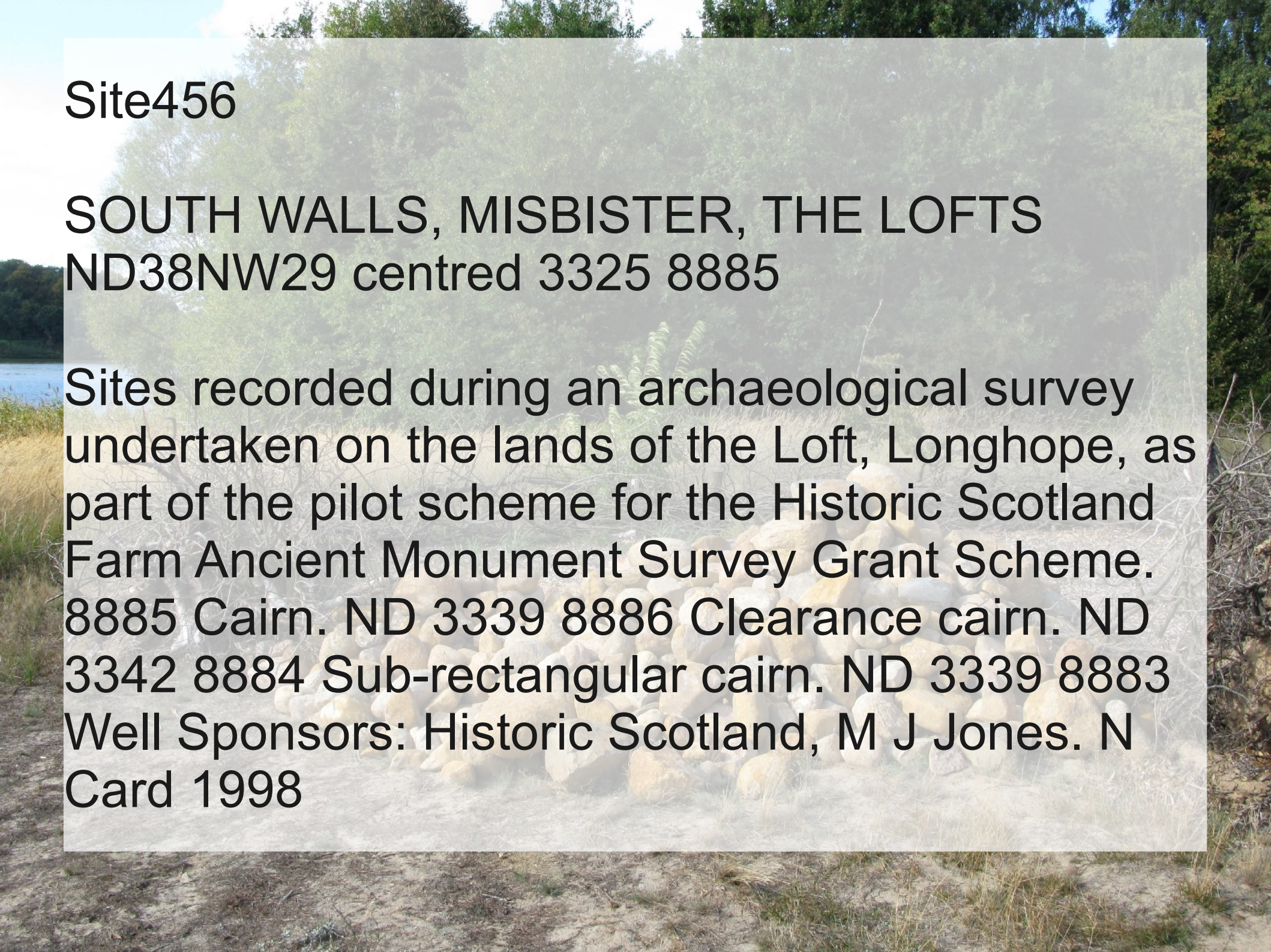



Figure 3: The text to RDF pipeline

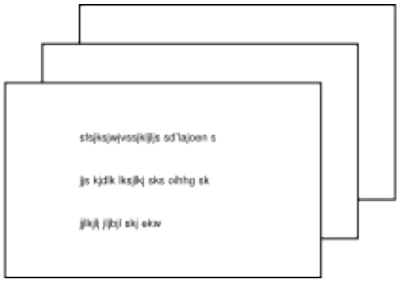


Site456

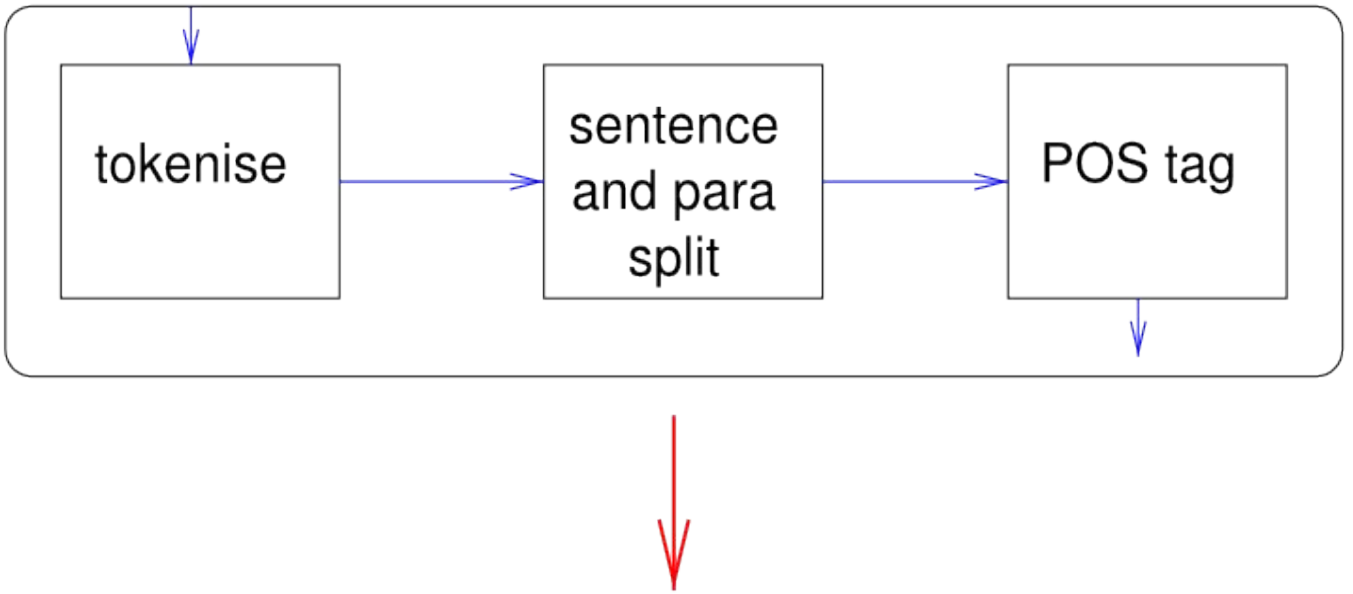
SOUTH WALLS, MISBISTER, THE LOFTS
ND38NW29 centred 3325 8885

Sites recorded during an archaeological survey undertaken on the lands of the Loft, Longhope, as part of the pilot scheme for the Historic Scotland Farm Ancient Monument Survey Grant Scheme. 8885 Cairn. ND 3339 8886 Clearance cairn. ND 3342 8884 Sub-rectangular cairn. ND 3339 8883 Well Sponsors: Historic Scotland, M J Jones. N Card 1998

***Text documents***



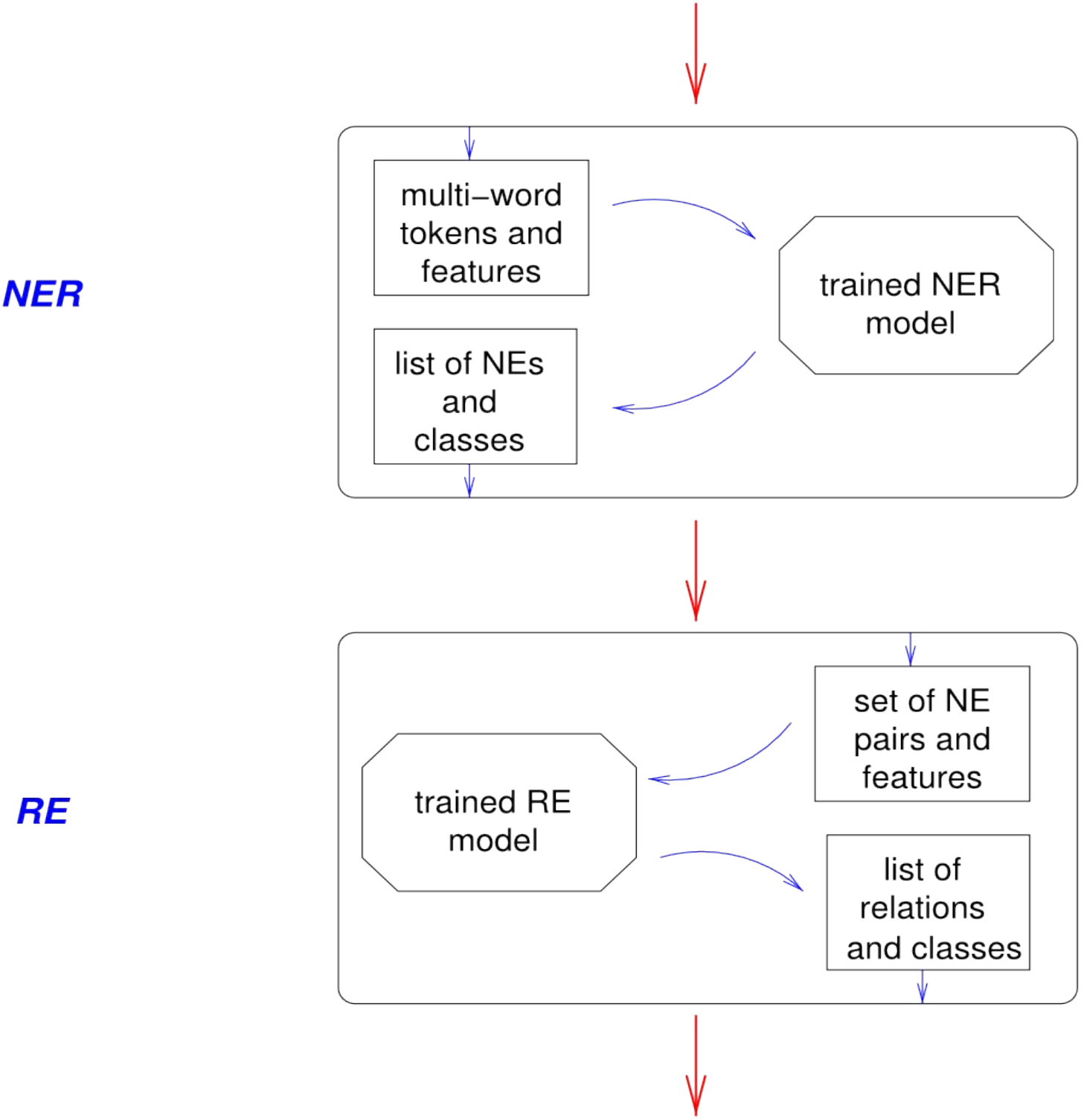
***Pre–processing***



{[SOUTH WALLS]} , {[MISBISTER]} , [[[[THE {LOFTS]]]]}

[ND38NW 29 centred 3325 8885]

Sites [[[[[[recorded]]]]]] during an [archaeological survey] undertaken on the lands of [[the {Loft]]} , [Longhope] , as part of the pilot scheme for the {[[[Historic {[Scotland]]]}} {Farm} {Ancient} {Monument} Survey Grant Scheme] . [ND 3311 8890] Two [small {cairns]} . [ND 3336 8889] {[Cairn]} . [ND 3339 8885] {[Cairn]} . [ND 3339 8886] {[Clearance cairn]} . [ND 3342 8884] [Sub-rectangular cairn]} . [ND 3339 8883] {[Well]} Sponsors : {[Historic {[Scotland]]}} , {[M J Jones]} . {{[[N Card]} {[1998]]


NER
multi–word tokens and features
trained NER model
list of NEs and classes
RE
set of NE pairs and features
trained RE model
list of relations and classes

# Named Entity Recognition


| | | | | | |
|---|---|---|---|---|---|
| *eventLocation* | were_found | unst | *cls1=event* | *cls2=place* | *wdsep=+2 ...* |
| *eventAgent* | were_found | a_t_cluness | *cls1=event* | *cls2=persname* | *wdsep=+5 ...* |
| *eventPatient* | were_found | steatite_dish | *cls1=event* | *cls2=artefact* | *wdsep=+9 ...* |
| *O* | unst | a_t_cluness | *cls1=place* | *cls2=persname* | *wdsep=+9 ...* |
| *O* | unst | steatite_dish | *cls1=place* | *cls2=artefact* | *wdsep=+9 ...* |
| *O* | a_t_cluness | steatite_dish | *cls1=persname* | *cls2=artefact* | *wdsep=+9 ...* |

# Relation Extraction

| | Form | Description |
|---|---|---|
| 1 | ne1=... | first NE string (concatenated using "_") |
| 2 | ne2=... | second NE string |
| 3 | cls1=... | first NE type |
| 4 | cls2=... | second NE type |
| 5 | wdsep=∘ $n$ | distance between NEs (+ve or -ve) |
| 6 | insent=y or n | both NEs in same sentence? |
| 7 | inpara=y or n | both NEs in same paragraph? |
| 8 | lastNEwdsame=y or n | normalised last token matches? |
| 9 | prevpos1=... | POS tag of token preceeding first NE |
| 10 | prevpos2=... | POS tag of token preceeding second NE |
| 11 | 1begsent=y or n | first NE is at beginning of a sentence |
| 12 | 2begsent=y or n | second NE is at beginning of a sentence |
| 13 | 1endsent=y or n | first NE is at end of a sentence |
| 14 | 2endsent=y or n | second NE is at end of a sentence |
| 15 | nest=n, 1in2 or 2in1 | one NE is nested within the other |
| 16 | neBetw=$n$ | number of NEs between this pair |
| 17 | verb=... | if insent=y, (first) verb between NEs; else "none" |

Table 1: Textual features used for building RE model.

**site456**

{[SOUTH WALLS]} , {[MISBISTER]} , [[[[THE {LOFTS]]]]}

[ND38NW 29 centred 3325 8885]

Sites [[[[[[recorded]]]]]] during an [archaeological survey] undertaken on the lands of [[the {Loft]]} , [Longhope] , as part of the pilot scheme for the {[[[Historic {[Scotland]]]}} {Farm} {Ancient} {Monument} Survey Grant Scheme] . [ND 3311 8890] Two [small {cairns]} . [ND 3336 8889] {[Cairn]} . [ND 3339 8885] {[Cairn]} . [ND 3339 8886] {[Clearance cairn]} . [ND 3342 8884] [Sub-rectangular cairn]} . [ND 3339 8883] {[Well]} Sponsors : {[Historic {[Scotland]]}} , {[M J Jones]} . {{[[N Card]} {[1998]]

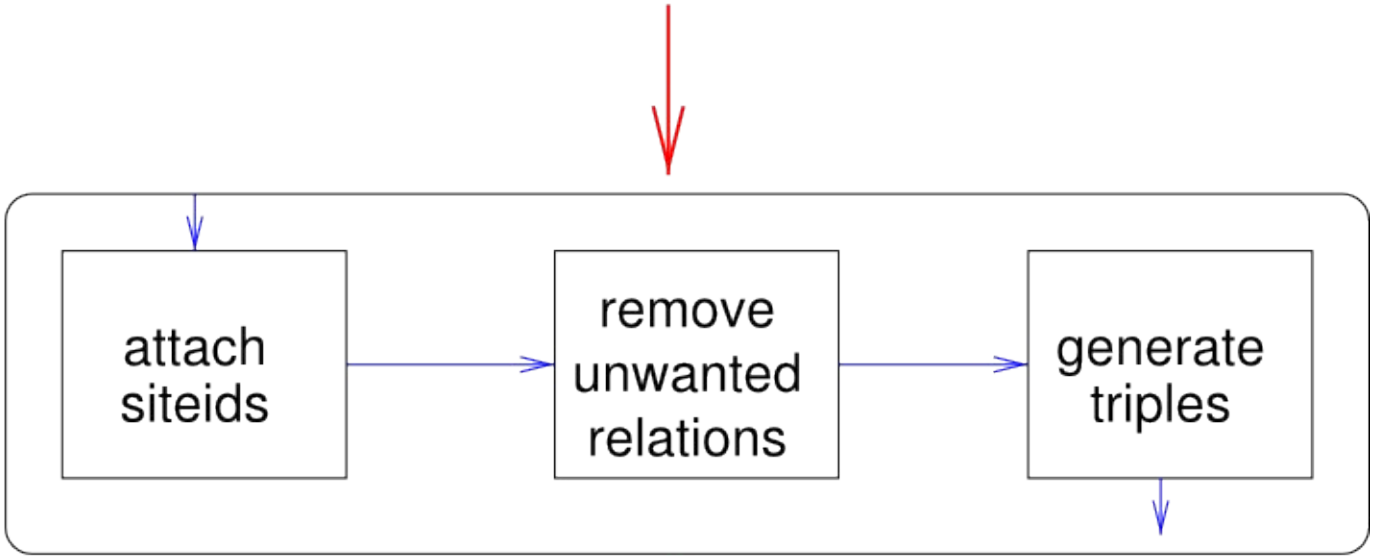
event

eventPatient

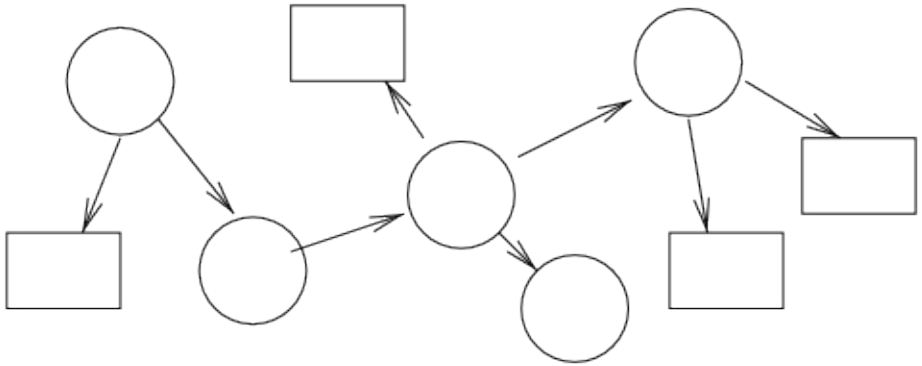
eventPlace

| |
|---|
| ADDRESS |
| PLACE |
| SITENAME |
| DATE |
| PERIOD |
| EVENT |
| ORG |
| PERSNAME |
| ROLE |
| SITETYPE |
| ARTEFACT |

***RDF translation***




***Graph of triples***



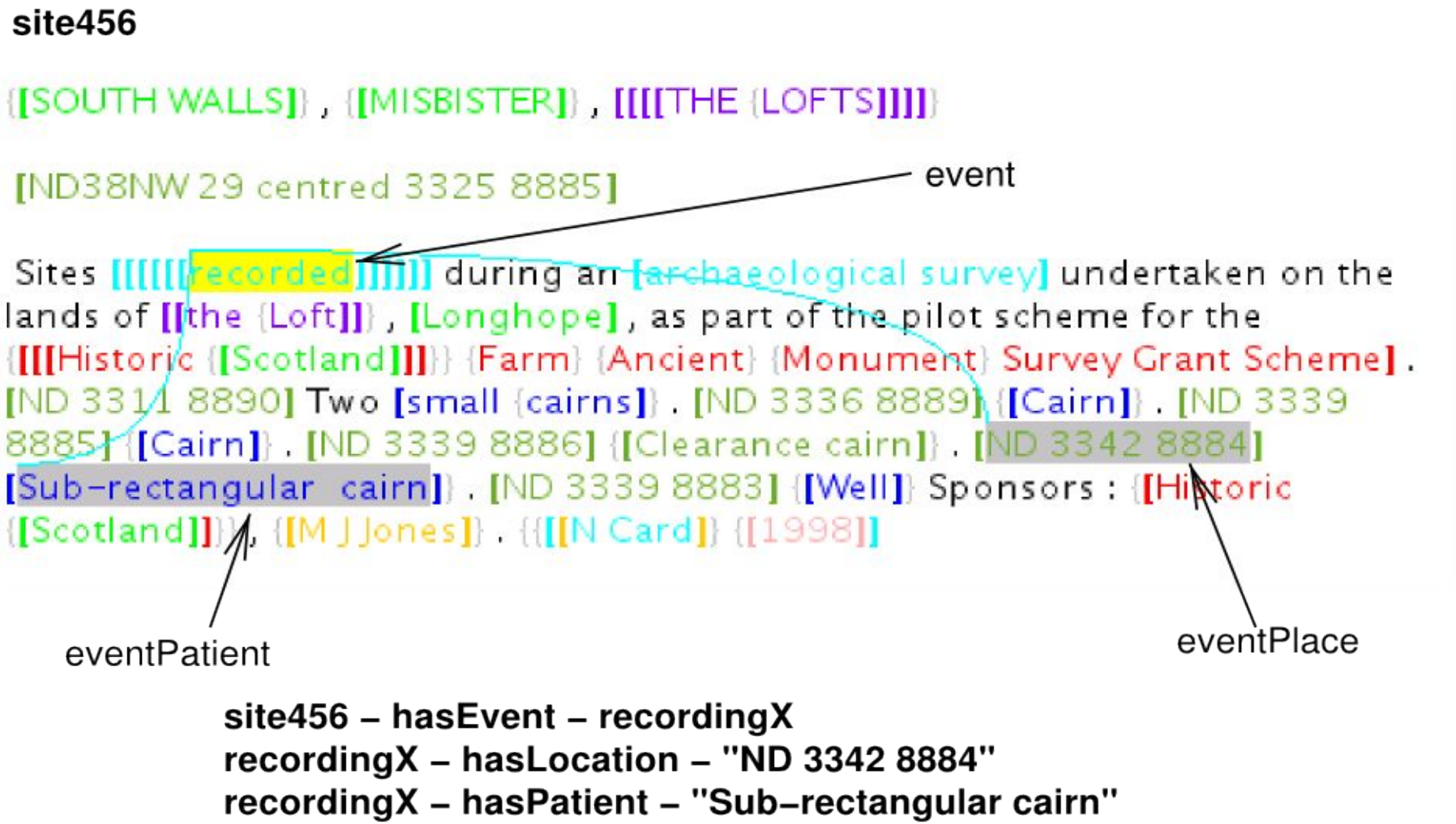
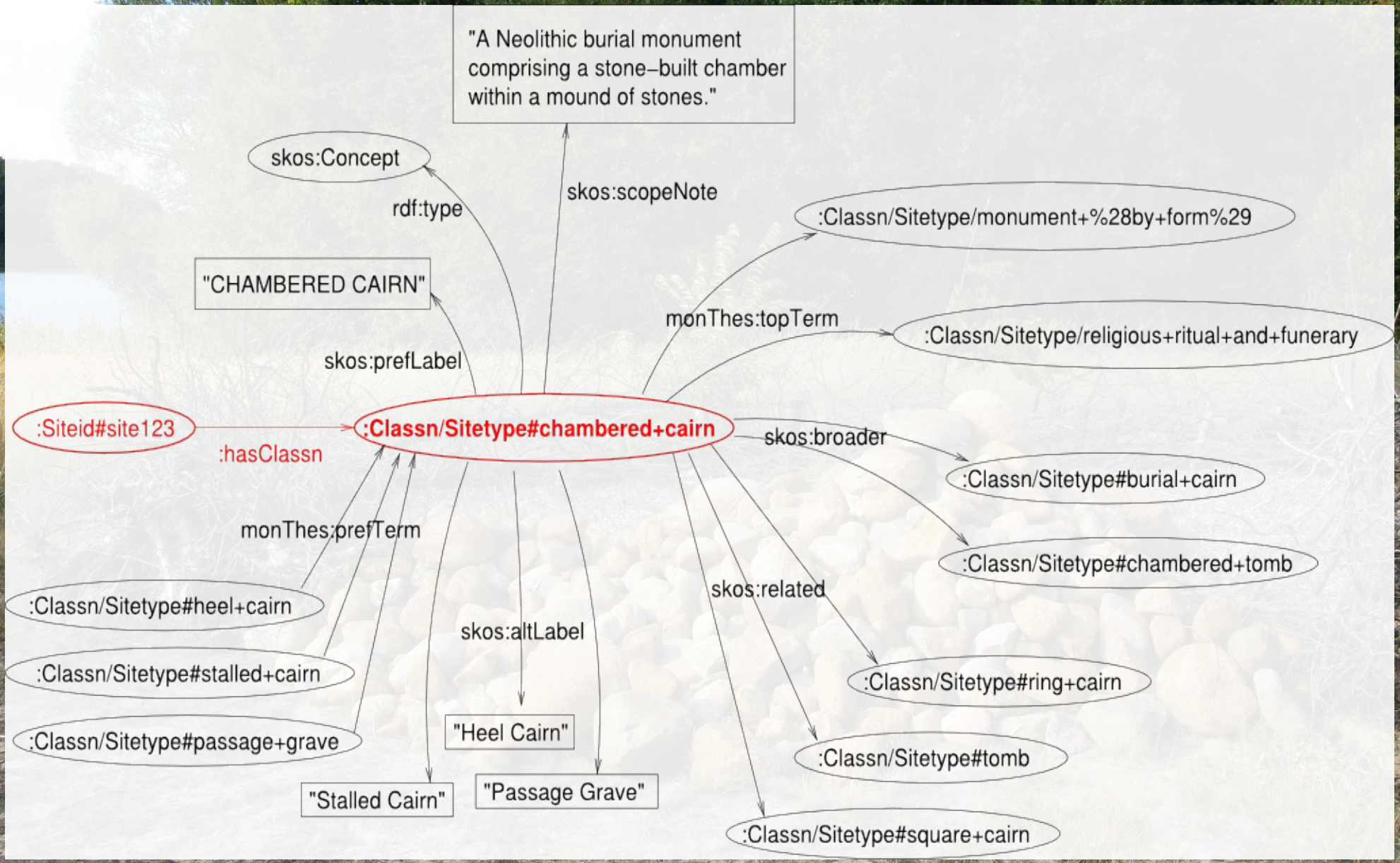
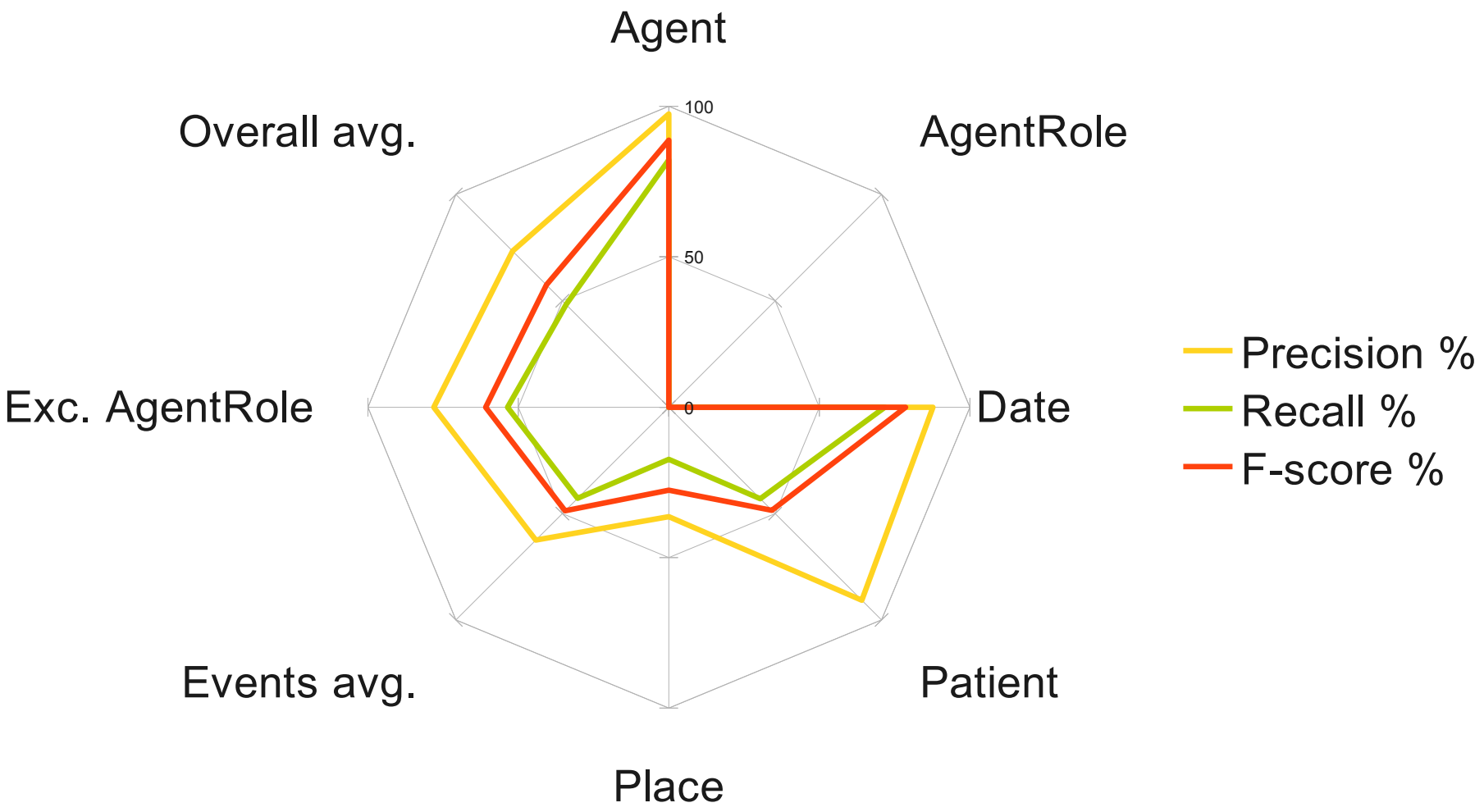



Figure 5: Mapping relations to RDF.


"A Neolithic burial monument comprising a stone–built chamber within a mound of stones."
skos:Concept
rdf:type
skos:scopeNote
:Classn/Sitetype/monument+%28by+form%29
"CHAMBERED CAIRN"
monThes:topTerm
:Classn/Sitetype/religious+ritual+and+funerary
skos:prefLabel
:Siteid#site123
:Classn/Sitetype#chambered+cairn
skos:broader
:hasClassn
:Classn/Sitetype#burial+cairn
monThes:prefTerm
:Classn/Sitetype#chambered+tomb
skos:related
:Classn/Sitetype#heel+cairn
skos:altLabel
:Classn/Sitetype#stalled+cairn
:Classn/Sitetype#ring+cairn
:Classn/Sitetype#passage+grave
"Heel Cairn"
:Classn/Sitetype#tomb
"Stalled Cairn"
"Passage Grave"
:Classn/Sitetype#square+cairn

# Results: whole pipeline NER + RE




Questions?

# Image sources

- The standing stones of Callanish on the Isle of Lewis, Scotland.

  - http://commons.wikimedia.org/wiki/File:Callanish-circle.jpg
- Clearance cairn at landscape conservation area Königswald (Potsdam).

  - http://commons.wikimedia.org/wiki/File:Lesesteinhaufen,_K%C3%B6nigswald,_Potsdam.jpg
- Memorial cairn at the Isandlwana hill in Kwazulu Natal, South Africa.

  - http://en.wikipedia.org/wiki/File:Isandlwanamassgrave.JPG