



Leibniz-Institut für Informationsinfrastruktur

Temporal Role Annotation for Named Entities

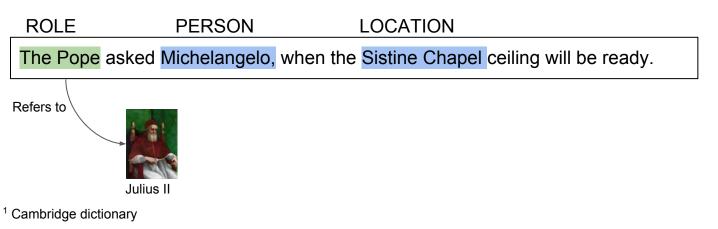
Maria Koutraki, Farshad Bakhshandegan-Moghaddam, Harald Sack Vienna, 13.9.2018

What is a role?

"the position or purpose that someone or something has in a situation, organization, society, or relationship"¹

- A role is a position
- Anything (in particular named entities) can take over a role

A role is not a named entity, refers to a named entity!





What is a role?

"the position or purpose that someone or something has in a situation, organization, society, or relationship" ¹

- A role is a position
- Anything (in particular *named entities*) can take over a role

A role is not a named entity, refers to a named entity!

ROLEPERSONLOCATIONThe PopeaskedMichelangelo,when theSistine Chapelceiling will be ready.

Categories:

- Permanent roles (e.g. human, mother, ...)
- **Temporal restricted roles** (e.g. CEO of a company, Pope, ...)

¹ Cambridge dictionary



Why (Temporal) Roles?

Behind the Pope's Embrace of Castro Shalis

The warmth and hospitality that pope Proncis showed in last week has buildfued many atholics - and for good reason he declator went to Rome for a last week has been before the state of the second state builting their encounter Gastor

the faith with a quit

<text>

gels. With God all things are possi-ble. Sut I suspect that this papal rapprochement with Castro has The Holy Father is a native of

cally defined by nationalism, so-challsim, corporatism and anti-americanism. It wouldn't be surprising to learn that this infla-cince this views toward the U.S. and the island 90 miles from its abore. When the Caban dictatorship

<text><text><text><text><text>

economies. There is another more planai-ble explanation for why the pope shows diadam in fas exploretation for "a crude and naive trust in the goadness of those weishing economic power and in the economic power and in the sacralized workings of the pre-valing economic system." It lies

In an Argentine sense of cultural respectority over the mixing-gradue finish in the state to pretect it. Menican Initiation and a state Reader that the state of an interlac-formate traces that to an interlac-tive the Opensiah defeat in the spentish American way. Therefore, he dies in his 2011 back Feed desencer's fortide the forcemans

Speculation runs from a Trojan horse plan to Latin American antipathy of the U.S.

<text><text>

gie between the North and the South. Many Latin American torship but they lowthe U.S. affluence and power. They know that a full-holwar, collapse of Cubs would likely bring back the would likely bring back the second likely provided the second are the shutts que.

would likely bring hore the second second second second second second the the start space. The provide second second second second the second second second second second the second second second second second the second secon

Last week Nev. Gustavo Gu Last week New Gustavo Cu-tiererea, the Perculan who be perculan who ournalists the Perculan who was hack at the Vatican. He tod ournalists that the church never ing and praised Pope Presence views on powerty Me didn't vian powerty since policy makers three out his ideal. Moybe the Bergeomber trip to Cuba.









FIZ Karlsruhe Leibniz-Institut für Informationsinfrastruktur



More Motivation...

- Roles for entities in knowledge graphs are present for most entity types
- Roles represent pivotal information for an entity
- Roles are in abundance in unstructured textual resources
- Roles are highly sparse in existing KGs
- Roles are tied to specific temporal and contextual information



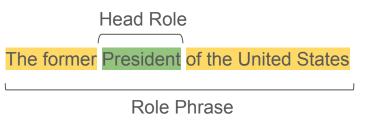
What are temporal restricted roles and how can we annotate them?



What is a *temporal* role?

- Roles that specify one or a few particular entities at a specific point in time
- A function which for a given a *temporal context* returns a *named entity* (or a few)!

- Head Role: Identifies the most general variant of a role with potentially high ambiguity
- **Role Phrase:** Specifies the role with sufficient details for the successful disambiguation.





Objective - Temporal Role Annotation

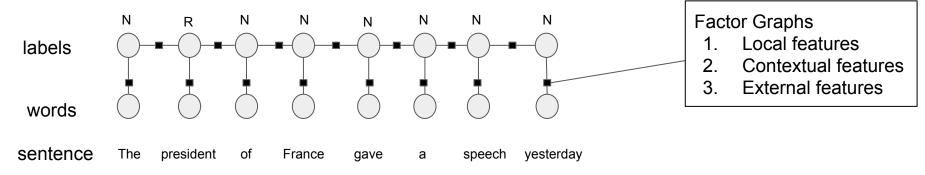
Goal: Detect and annotate in an automatic way

- temporal roles and
- role phrases in natural language text



Approach - Head Role Annotator

- Automated approach \rightarrow sequence classification
 - The roles are influenced by the preceding and succeeding words in a sentence
 - The relation between different segments in a sentence has impact in categorization
- Linear-chain CRF model:





Approach - Local Features

Take into account only the information from the individual words

id	Feature
f ₁	w
f ₂	POS(w _i)
f ₃	NER(w _i)
f ₄	startWithCapital(w _i)
f ₅	fullyInCapital(w _i)
f ₆	startOfSentence(w _i)
f ₇	Lemma(w _i)

Example: The president of France gave a speech yesterday

Feature Values for i=1: f_1 ="The", f_2 ="DT", f_3 ="null", f_4 ="true", f_5 ="false", f_6 ="true", f_7 ="the"

Feature Values for i=2: f₁="president", f₂ ="noun", f₃="null", f₄="false", f₅="false", f₆="false", f₇="president"



Capture the dependencies of a word w_i and its surrounding words $w_{i+2} = w_{i-2}$

id	Feature
f ₈₋ f ₁₁	$W_{i-2}^{}$, $W_{i-1}^{}$, $W_{i+1}^{}$, $W_{i+2}^{}$
f ₁₂₋ f ₁₅	$\begin{array}{l} POS(w_{i-2}), POS(w_{i-1}), \\ POS(w_{i+1}), POS(w_{i+2}) \end{array}$
f ₁₆₋ f ₁₉	$\begin{array}{l} {\sf NER}(w_{_{i-2}}), {\sf NER}(w_{_{i-1}}), \\ {\sf NER}(w_{_{i+1}}), {\sf NER}(w_{_{i+2}}) \end{array}$

Example: The king in chess is the most important piece

Feature Values for i=2: f_8 ="null", f_9 ="The", f_{10} ="in", f_{11} ="chess"



Capture the dependencies of a word w_i and its surrounding words $w_{i+2} = w_{i-2}$

id	Feature
f ₈₋ f ₁₁	$W_{i-2}^{}$, $W_{i-1}^{}$, $W_{i+1}^{}$, $W_{i+2}^{}$
f ₁₂₋ f ₁₅	$\begin{array}{l} POS(w_{_{i\!-\!2}}), POS(w_{_{i\!-\!1}}), \\ POS(w_{_{i\!+\!1}}), POS(w_{_{i\!+\!2}}) \end{array}$
f ₁₆₋ f ₁₉	$\begin{array}{l} {\sf NER}(w_{_{i-2}}), {\sf NER}(w_{_{i-1}}), \\ {\sf NER}(w_{_{i+1}}), {\sf NER}(w_{_{i+2}}) \end{array}$

Exploit the commonalities in the language structure!

Example: The <u>president</u> of France gave a speech yesterday

Feature Values for i=2: f₁₂="null", f₁₃ ="DT", f₁₄="IN", f₁₅="NNP"



Approach - External Features

Use knowledge from external sources

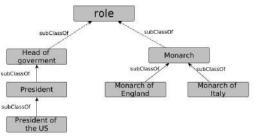
id	Feature			
f ₂₀	isInDic(w _i)			
f ₂₁	LocalGrammar(w _i)			

Boolean flag indicating if a token is part of a dictionary



Approach - External Features - Dictionary

- No existing dictionary dedicated to temporal roles
- Automatically construct a dictionary using knowledge graphs (Wikidata)
- Wikidata class *role(Q214339)* subtree:



 To ensure the temporal aspect of roles consider time-dependent properties (e.g. "replaces", "replaced by")

SELECT DISTINCT ?role ?roleLabel WHERE {
?role instanceOf*|subClassOf* wd:Q214339 .
?role label ?roleLabel.
?person positionHeld ?roleStatement.
?roleStatement positionHeld ?role.
?roleStatement replaces|replacedby ?differentRoleHolder.



query:

Approach - External Features

Use knowledge from external sources

id	Feature			
f ₂₀	isInDic(w _i)			
f ₂₁	LocalGrammar(w _i)			

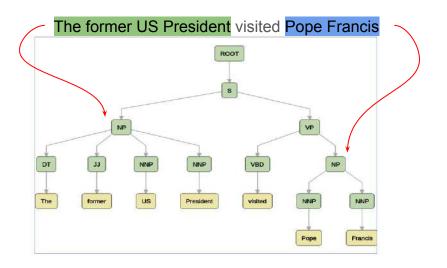
• Boolean flag indicating if a token matches to a local grammar

Local Grammar	Samples	
ROLE IN (POS) LOCATION (NER)	President of Iran, King of Egypt	
ROLE PERSON (NER)	President Obama, Pope Francis	
DT(POS) ROLE	The King, The Pope	
ROLE IN (POS) ORGANIZATION (NER)	CEO of Apple	
ORGANIZATION (NER) ROLE	Google CEO	



Approach - Role Phrase Annotation

- Constituency Parse Tree
- Select nearest to the root Noun-Phrase





Experimental Setup (1/2)

- Tested Roles:
 - President, Pope, Monarch, CEO

- Baselines:
 - Temporal Role Dictionary (>200 unique temporal role surfaces)
 - ANNIE job title gazetteer (>1.500 job titles)



Experimental Setup (2/2)

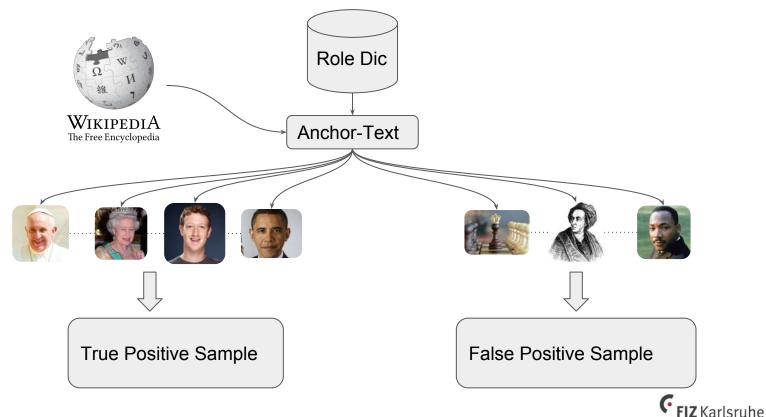
• Ground Truth:

- Manually created dataset
 - New York Times corpus
 - 200 sentences
 - 64% True Positive 36% False Positive
- Automatically generated dataset
 - Wikipedia articles

Sample	#tokens
True Positive	6M
False Positive	1.5M
True Negative	3.3M



Automatically generated GT Creation



19

Leibniz-Institut für Informationsinfrastruktur

Evaluation - Performance

• 5-fold cross validation over the automatically extracted dataset

Method	Р	R	F ₁
Temp Role Dictionary	0.82	1	0.9
CRF-based Role Annotator	0.94	0.97	0.95

• The dictionary filters out the true negative samples but not the samples with ambiguous surface forms which may refer to a person name or a role (e.g. Pope)



Evaluation - Robustness

• Trained on the automatically extracted dataset (wikipedia) tested on the manually created dataset (New York Times)

Method	Р	R	F ₁
Temp Role Dictionary	0.66	1	0.79
ANNIE Gazetteer	0.55	0.72	0.62
CRF-based Role Annotator	0.80	0.93	0.86

- Dictionary: Cannot differentiate the "Alexander Pope" and "Pope Francis"
- Role Annotator: learns patterns like "<PERSON> Pope" → Not a Role



Evaluation - Feature Ablation

• 5-fold cross validation over the automatically extracted dataset (Wikipedia)

Feature Group	Р	R	F ₁
Local Features	0.90	0.95	0.92
Contextual Features	0.72	0.27	0.39
External Features	0.82	1	0.90
All Features	0.94	0.97	0.95
All Features (25% Dictionary size reduction)	0.94	0.92	0.93



Conclusion and Future Steps

- Definition of the task of temporal role annotation
- Proposed a structured prediction approach for the annotation of temporal roles
- Constructed two high quality ground-truth datasets for this task

• Next Steps

- Experiment with more Role categories
- Enrich the ground truth
- Experiment with more features
- Continue with the subsequent step of disambiguating the temporal roles to the corresponding named entities



Thank You!

Code and Data: <u>https://github.com/ISE-AIFB/RoleTagger</u>

Homepage: https://www.fiz-karlsruhe.de/en/forschung/information-service-engineering/staff-ise/dr-maria-koutraki.html

Email: maria.koutraki@fiz-karlsruhe.de, maria.koutraki@kit.edu

Twitter account: <u>@mairyl0u</u>





Related Work

- NER
 - NER approached are optimized to label surface forms that link to a NE class (PERSON, ORGANIZATION, PLACE...)
- NED/EL
- Dictionaries and Gazetteers [1]
 - Domain dependent
 - Manual effort
 - Do not scale
 - Suffer in terms of recall

[1] Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V., 2002. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. ACL'02

