Fine-grained Named Entity Recognition in Legal Documents

Elena Leitner, Georg Rehm, Julian Moreno-Schneider

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH
Alt-Moabit 91c, 10559 Berlin, Germany

SEMANTiCS
Karlsruhe 2019

LYNX: BUILDING THE LEGAL KNOWLEDGE GRAPH FOR SMART COMPLIANCE SERVICES IN MULTILINGUAL EUROPE

Grant Agreement number: 780602
ICT14-2016-2017 / Innovation Action / Horizon 2020
Pillar: Industrial Leadership
Work Programme Year: H2020-2016-2017
Work Programme Part: Information and Communication Technology

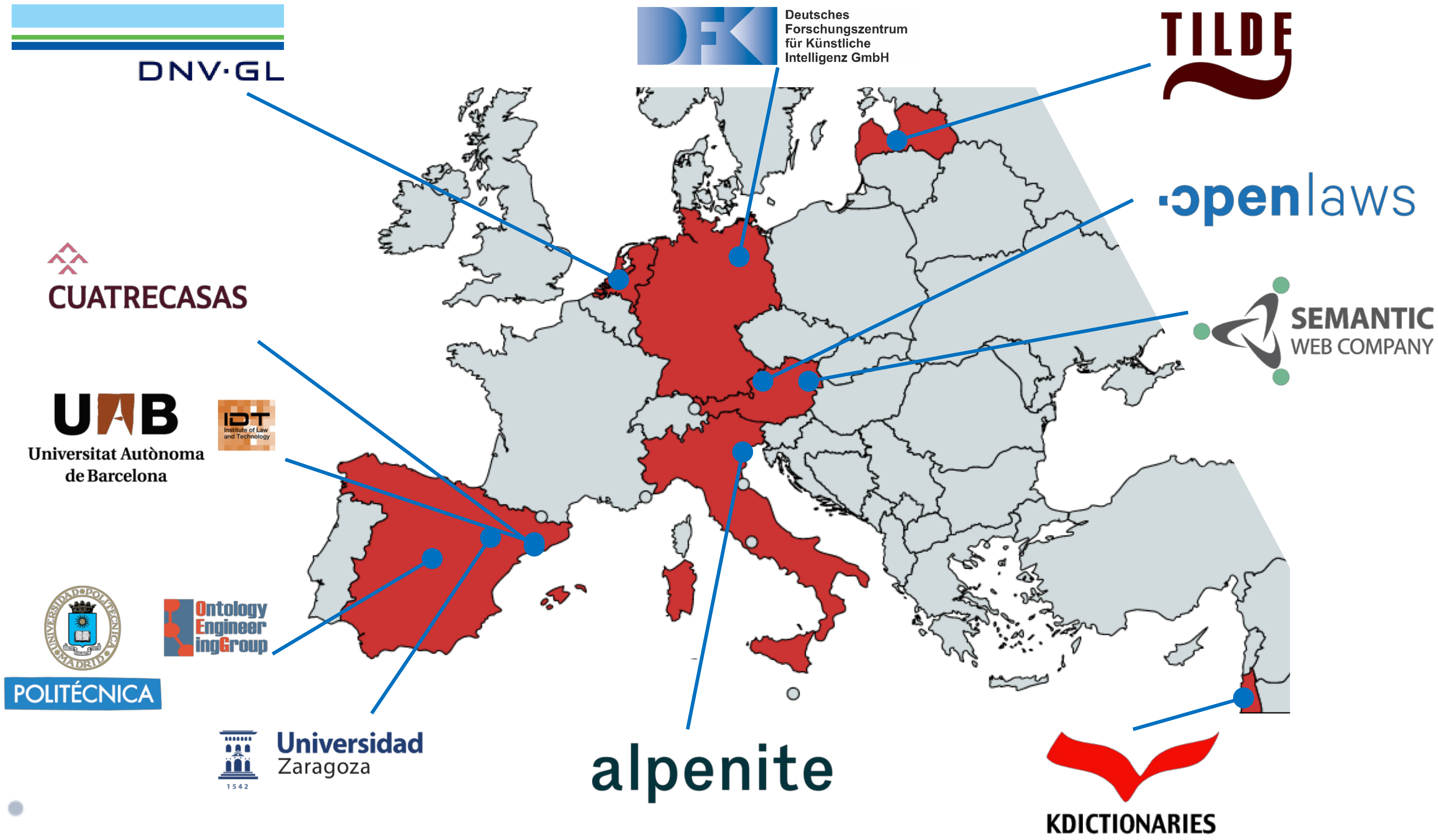
Big Data PPP: cross-sectorial and cross-lingual data integration and experimentation

Duration: 36 months
Starting date: December 1, 2017
Project Cost: €3,638,065.00
EU Contribution: €2,959,247.52

European Commission | Horizon 2020 European Union funding for Research & Innovation

BDV BIG DATA VALUE ASSOCIATION

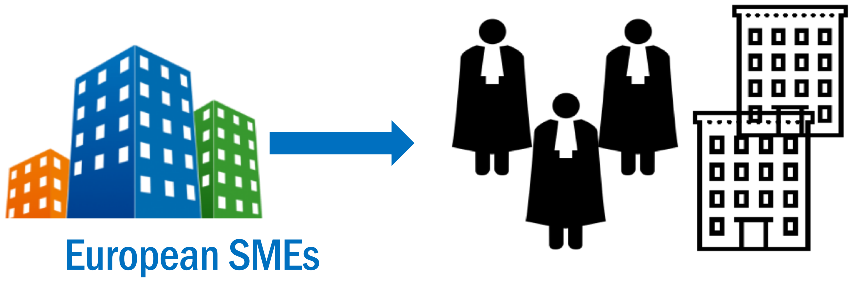
WHO & WHERE

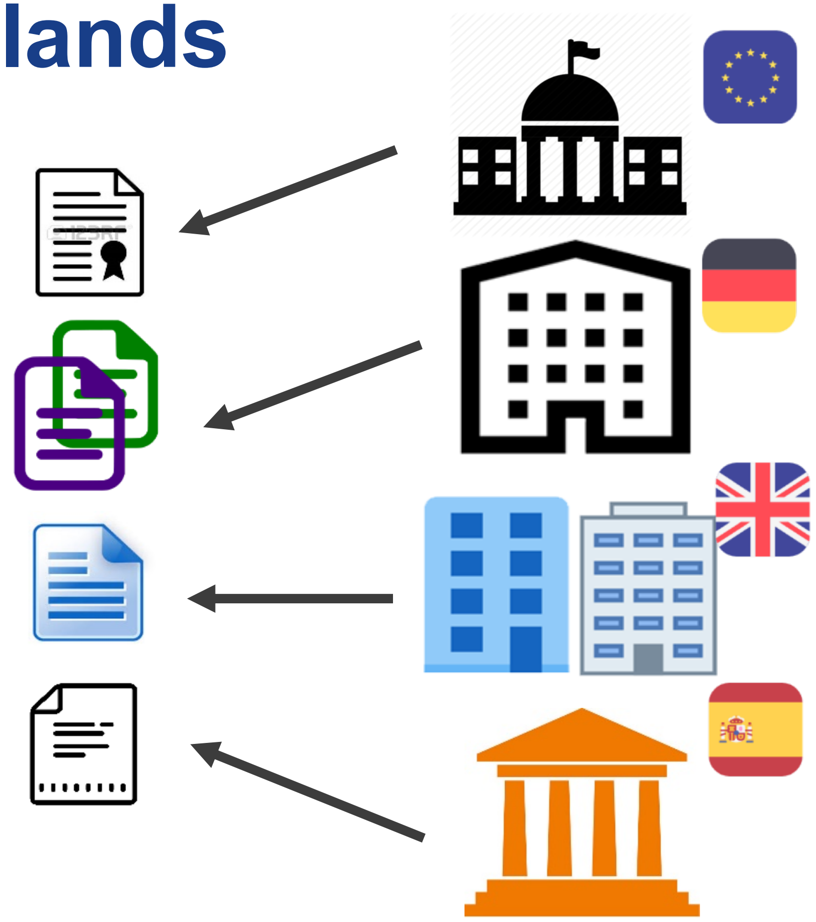


DNV·GL
Deutsches Forschungszentrum für Künstliche Intelligenz GmbH
TILDE
openlaws
SEMANTIC WEB COMPANY
CUATRECASAS
UAB Universitat Autònoma de Barcelona
IDT Institute of Law and Technology
POLITÉCNICA
Ontology Engineering Group
Universidad Zaragoza
alpenite
KDICTIONARIES

BACKGROUND

SMEs and large corporations face multiple constraints to trade in the EU and to localize their products and services, due to **legal silos** and **linguistic islands**





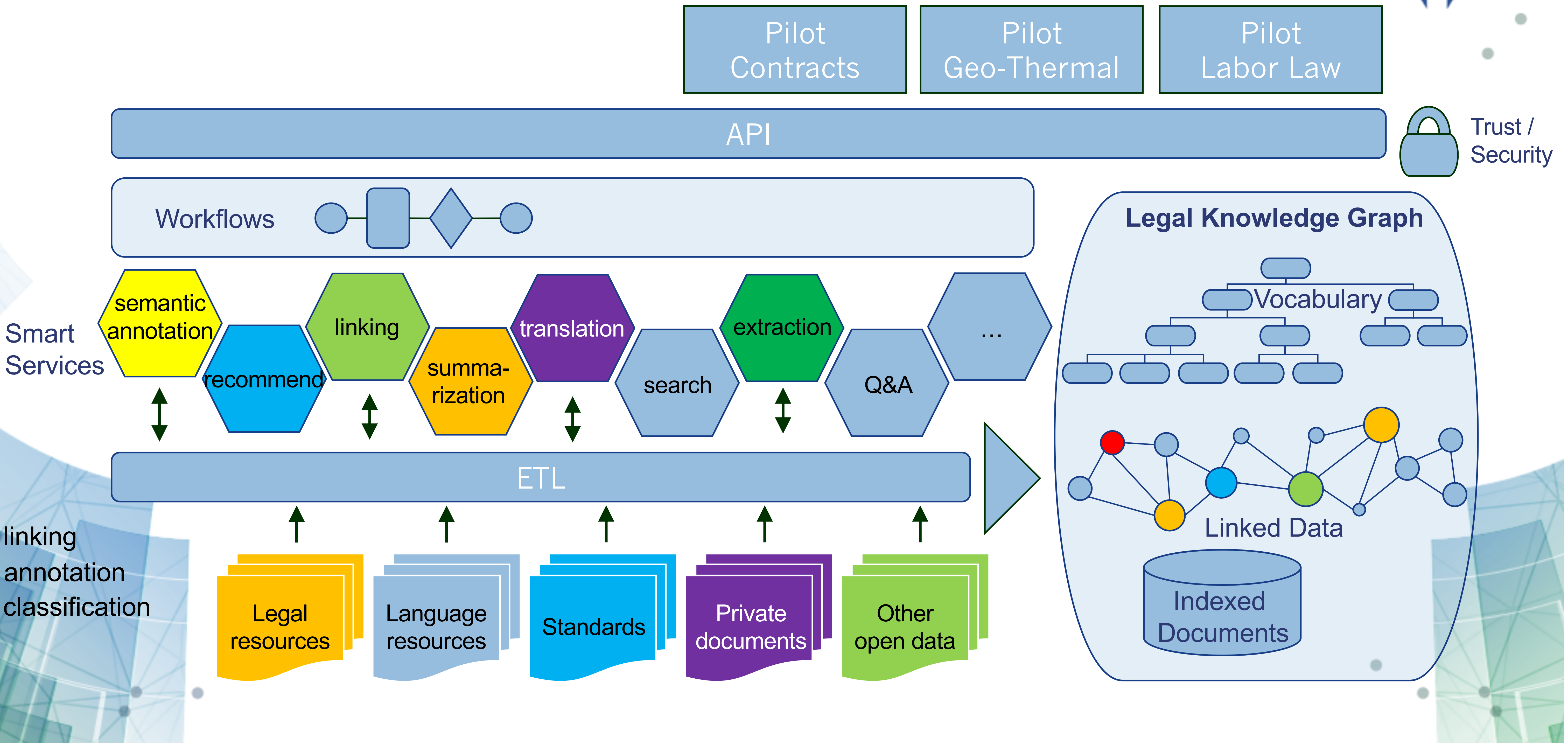
COMPANIES WISHING TO OPERATE IN A NEW MARKET MUST

- Comply with legislation (European, national, regional, local)
- Implement **different standards** (e.g., ISO, AENOR, DIN)
- Follow sector-specific practices

OBJECTIVES

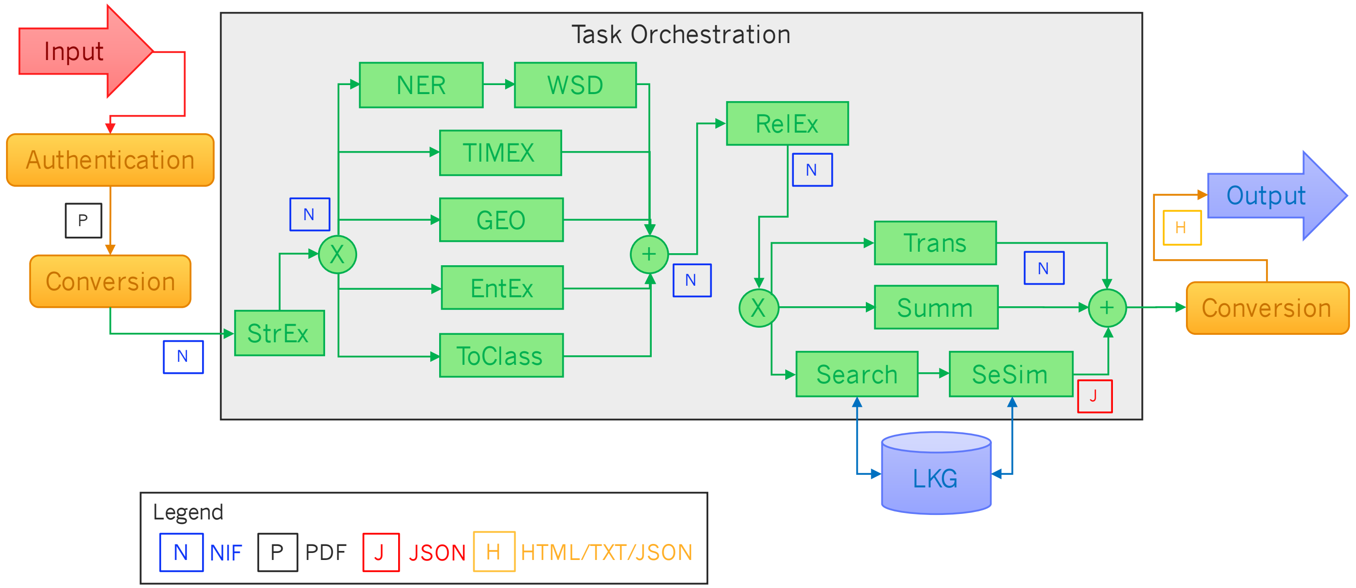
Create an ecosystem of smart cloud services to better manage compliance, based on a legal knowledge graph which integrates and links heterogeneous compliance data sources including legislation, case law, standards and other private data.

ENVISIONED SOLUTION

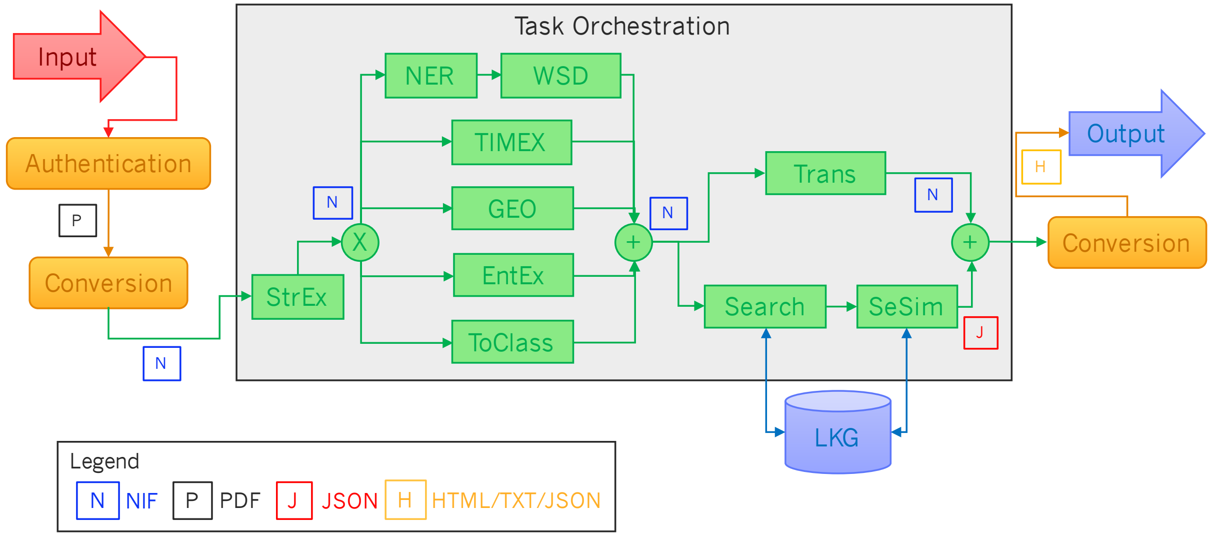


THREE PILOTS AND WORKFLOWS

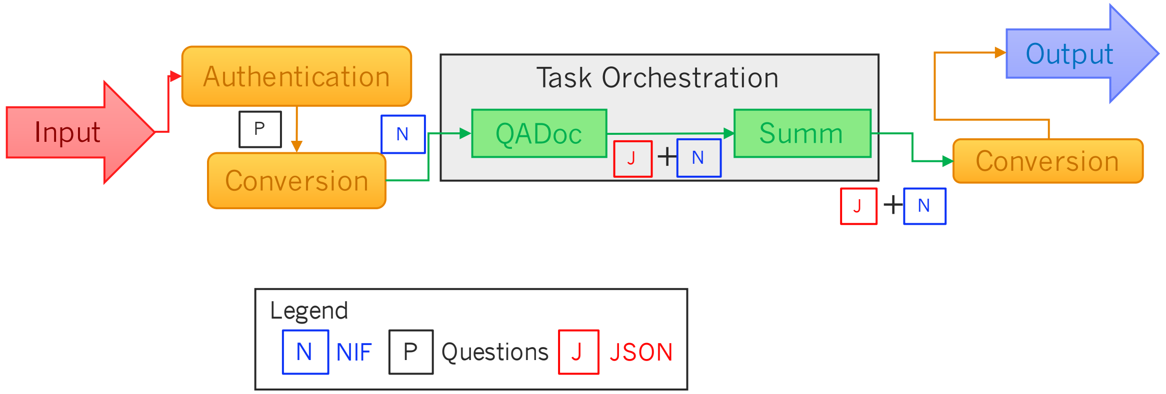
CONTRACT ANALYSIS WORKFLOW



GEOTHERMAL ENERGY WORKFLOW

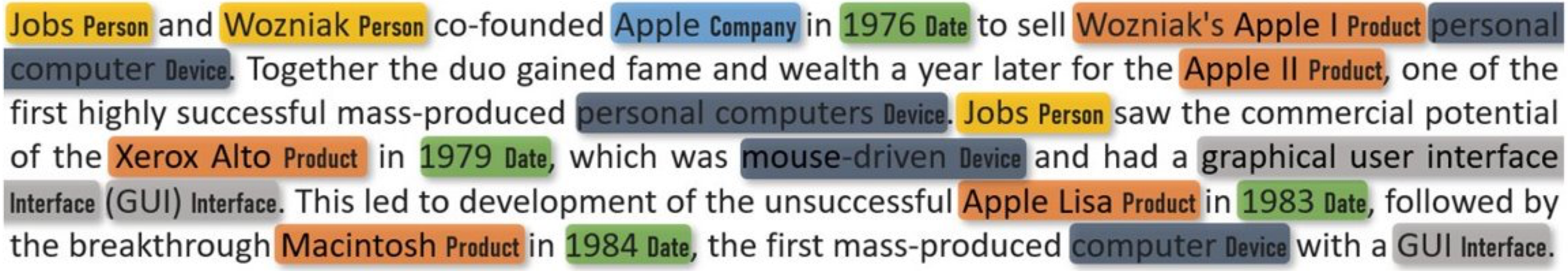


LABOUR LAW SEARCH WORKFLOW



Introduction

Named Entity Recognition (NER) is the automatic identification and classification of named entities (NEs) in text into predefined categories such as *person*, *organization*, *location*, etc.



Research Goal

Examine Named Entity Recognition in German legal documents:

- Elaboration of corresponding semantic concepts
- Construction of a dataset
- Developing, evaluating and comparing state of the art models for NER

Related Work in Legal Domain

Authors	Source texts	New classes	Techniques	Results
Dozier et al. (2010)	• US case law • Depositions • Pleadings • Other	• *Jurisdiction* • *Court* • *Title* • *Document type* • *Judge*	• Lookups • Contextual rules • Statistical models	• 82–85 F1
Cardellino et al. (2017)	• Wikipedia • Decisions of the European Court of Human Rights (ECHR)	• In NERC: *document, abstraction, act* Granularity levels: • NER • NERC • LKIF • YAGO	• SVM • Stanford NER • NN	Wikipedia: • Stanford NER (LKIF) – 77 F1 • NN (NERC) – 86 F1 • NN (YAGO) – 69 F1 Decisions of ECHR: • Stanford NER (NERC) < 56 F1
Glaser et al. (2018)	• Court decisions • Contracts (for templated technique)	• *Reference*	• GermaNER (+ rule-based approaches, references recognizer) • DBpedia Spotlight • Templated	• GermaNER – 80 F1 • DBpedia Spotlight – 87 F1 • Templated – 92 F1

State of the Art

Authors	System type	Performance CoNLL 2003 (F1)	
		EN	DE
Tkachenko und Simanovsky (2012)	CRF	91.02	
Passos et al. (2014)	CRF	90.90	
Benikova et al. (2015) GermaNER	CRF		79.37
Huang et al. (2015)	BLSTM-CRF	90.10	
Lample et al. (2016)	BLSTM-CRF	90.94	78.76
Chiu & Nichols (2016)	BiLSTM-CNN	91.62	
Ma und Hovy (2016)	BLSTM-CNN-CRF	91.21	
Current state of the art approaches use language models			
Peters et al. (2018)	LM-CNN-LSTM-CRF	92.22	
Akbik et al. (2018)	LM-BiLSTM-CRF	93.09	

Challenges and Prerequisites

Research Challenges:

- No freely available datasets of legal documents (English or German)
- No established semantic categories of NER in the legal domain
- No established annotation guidelines

Research Prerequisites:

- Development of a typology of semantic categories for legal documents
- Development of annotation guidelines
- Annotation of dataset for machine learning experiments

Semantic Categories

- Legal documents are a rather unique category of texts.
- Occurrence of typical NEs (PER, LOC, ORG) is quite low.
- However, mentions of reference laws, decisions, and regulations are rather typical.

Requirements for the elaboration of the typology are:

- Must reflect typical entities
- Must concern relevant entities

Person PER
- Person PER
- Judge RR
- Lawyer AN

Location LOC
- Country LD
- City ST
- Street STR
- Area LDS

Organization ORG
- Organization ORG
- Institution INN
- Company UN
- Court GRT
- Brand MRK

Legal norm NRM
- Law GS
- Ordinance VO
- European legal norm EUN

Case-by-case regulation REG
- Regulation VS
- Contract VT

Court decision RS

Legal literature LIT

Semantic Categories: Examples

(1) Das Ablehnungsgesuch der Beschuldigten vom 1. April 2018 gegen die Vorsitzende Richterin am Bundesgerichtshof **GRT** Sost-Scheible **RR**, die Richterin am Bundesgerichtshof **GRT** Roggenbuck **RR** und die Richter am Bundesgerichtshof **GRT** Cierniak **RR** , Bender **RR** und Dr. Feilcke **RR** wird als unzulässig verworfen.

(2) Jedoch wird der Verkehr darin naheliegend den Namen eines der bekanntesten Flüsse Deutschlands **LD** erkennen, welcher als Seitenfluss des Rheins **LDS** durch Oberfranken **LDS**, Unterfranken **LDS** und Südhessen **LDS** fließt und bei Mainz **ST** in den Rhein **LDS** mündet.

(3) Der FC Bayern München **ORG** schloss den Beschwerdeführer ... aus dem Verein aus ...

(4) Die Landesregierung Rheinland-Pfalz **INN** hat von einer Stellungnahme abgesehen.

(5) … des US-amerikanischen Unternehmens Apple **UN** …

(6) Vorliegend stehen sich die Widerspruchsmarke Becker Mining **MRK** und die angegriffene Marke Becker **MRK** gegenüber.

(7) … unter der Firma C . . . AG **UN** ...

(8) … der ebenfalls beim Bundesgerichtshof **GRT** zugelassene Rechtsanwalt ... **AN** …

Person **PER**
Judge **RR**
Lawyer **AN**
Country **LD**
City **ST**
Street **STR**
Area **LDS**
Organization **ORG**
Institution **INN**
Company **UN**
Court **GRT**
Brand **MRK**
Law **GS**
Ordinance **VO**
European legal norm **EUN**
Regulation **VS**
Contract **VT**
Court decision **RS**
Legal literature **LIT**

Semantic Categories: Examples

(9) … § 14 Absatz 2 Satz 2 des Gesetzes über Teilzeitarbeit und befristete Arbeitsverträge (TzBfG) vom 21. Dezember 2000 (Bundesgesetzblatt Seite 1966), zuletzt geändert durch Gesetz vom 20. Dezember 2011 (Bundesgesetzblatt I Seite 2854) **GS**, ist nach Maßgabe der Gründe mit dem Grundgesetz **GS** vereinbar.

(10) Mit der Neuregelung in § 35 Abs. 6 StVO **VO** …

(11) … insbesondere durch die Richtlinien zur Bewertung des Grundvermögens – BewRGr – vom 19. September 1966 (BStBl I, S. 890) **VS**.

(12) Auf das Arbeitsverhältnis der Parteien fand der Manteltarifvertrag für die Beschäftigten der Mitglieder der TGAOK **VT** (BAT/AOK-Neu **VT**) vom 7. August 2003 Anwendung.

(13) … (stRspr; vgl zB BVerfGE 62, 1, 45 **RS**; BVerfGE 119, 96, 179 **RS**; BSG SozR 4 – 2500 § 62 Nr 8 RdNr 20 f **RS**; Hauck/Wiegand, KrV 2016, 1, 4 **LIT**).

Person **PER**
Judge **RR**
Lawyer **AN**
Country **LD**
City **ST**
Street **STR**
Area **LDS**
Organization **ORG**
Institution **INN**
Company **UN**
Court **GRT**
Brand **MRK**
Law **GS**
Ordinance **VO**
European legal norm **EUN**
Regulation **VS**
Contract **VT**
Court decision **RS**
Legal literature **LIT**

Dataset

- 750 German court decisions (*Rechtsprechung im Internet*)
  https://www.rechtsprechung-im-internet.de
- 7 coarse-grained classes
- 19 fine-grained classes
- 66,723 sentences
- 2,157,048 tokens
- 53,632 entities annotated manually
- 19% annotations (per-token basis)

Coarse-grained classes			#	%	Fine-grained classes			#	%
1	**PER**	*Person*	3,377	6.30	1	**PER**	*Person*	1,747	3.26
					2	**RR**	*Judge*	1,519	2.83
					3	**AN**	*Lawyer*	111	0.21
2	**LOC**	*Location*	2,468	4.60	4	**LD**	*Country*	1,429	2.66
					5	**ST**	*City*	705	1.31
					6	**STR**	*Street*	136	0.25
					7	**LDS**	*Area*	198	0.37
3	**ORG**	*Organization*	7,915	14.76	8	**ORG**	*Organization*	1,166	2.17
					9	**UN**	*Company*	1,058	1.97
					10	**INN**	*Institution*	2,196	4.09
					11	**GRT**	*Court*	3,212	5.99
					12	**MRK**	*Brand*	283	0.53
4	**NRM**	*Legal norm*	20,816	38.81	13	**GS**	*Law*	18,520	34.53
					14	**VO**	*Ordinance*	797	1.49
					15	**EUN**	*EU legal norm*	1,499	2.79
5	**REG**	*Case-by-case regulation*	3,470	6.47	16	**VS**	*Regulation*	607	1.13
					17	**VT**	*Contract*	2,863	5.34
6	**RS**	*Court decision*	12,580	23.46	18	**RS**	*Court decision*	12,580	23.46
7	**LIT**	*Legal literature*	3,006	5.60	19	**LIT**	*Legal literature*	3,006	5.6
		Total	53,632	100			**Total**	53,632	100

Gericht:	BVerfG 2. Senat	**Normen:**	§ 24 S 2 BVerfGG, § 48 BVerfGG, § 26 Abs 3 S 3 EuWG
Entscheidungs-datum:	26.04.2018		
Aktenzeichen:	2 BvC 6/15		
ECLI:	ECLI:DE:BVerfG:2018:cs20180426.2bvc000615		
Dokumenttyp:	Beschluss		

Erledigung bzw Verwerfung (a-limine-Abweisung) einer Wahlprüfungsbeschwerde ohne weitere Begründung

Tenor

Die Wahlprüfungsbeschwerde der Beschwerdeführerin zu 1. ist durch ihren Tod erledigt.

Im Übrigen wird die Wahlprüfungsbeschwerde verworfen.

Gründe

1 Die Wahlprüfungsbeschwerde der Beschwerdeführerin zu 1. hat sich durch ihren Tod erledigt. Es kann dahinstehen, ob eine Fortführung der Wahlprüfungsbeschwerde durch einen Rechtsnachfolger zulässig wäre, da der Bevollmächtigte der Beschwerdeführerin zu 1. eine solche Person nicht benannt hat. Unter diesen Umständen ist lediglich auszusprechen, dass sich das Verfahren durch den Tod der Beschwerdeführerin zu 1. erledigt hat (vgl. BVerfGE 109, 279 <304>).

2 Im Übrigen bleibt der Wahlprüfungsbeschwerde aus den in dem Schreiben des Berichterstatters vom 8. Februar 2018 genannten Gründen der Erfolg versagt. Gemäß § 26 Abs. 3 Satz 3 EuWG in Verbindung mit § 24 Satz 2 BVerfGG wird von einer weiteren Begründung abgesehen.

Gericht:	BAG 6. Senat
Entscheidungs-datum:	22.03.2018
Aktenzeichen:	6 AZR 30/17
ECLI:	ECLI:DE:BAG:2018:220318.U.6AZR30.17.0
Dokumenttyp:	Urteil

Verfahrensgang

vorgehend ArbG Frankfurt, 2. Februar 2016, Az: 24 Ca 6460/15, Urteil
vorgehend Hessisches Landesarbeitsgericht, 28. November 2016, Az: 16 Sa 262/16, Urteil

Tenor

1. Auf die Revision der Beklagten wird das Urteil des Hessischen Landesarbeitsgerichts vom 28. November 2016 - 16 Sa 262/16 - aufgehoben.

2. Auf die Berufung der Beklagten wird das Urteil des Arbeitsgerichts Frankfurt am Main vom 2. Februar 2016 - 24 Ca 6460/15 - abgeändert.

Die Klage wird abgewiesen.

3. Der Kläger hat die Kosten des Rechtsstreits zu tragen.

Sonstiger Langtext

1 Die Parteien haben auf Tatbestand und Entscheidungsgründe verzichtet *(§ 313a Abs. 1 ZPO)*.

Fischermeier Krumbiegel Heinkel
Wollensak Kohout

CRF Training

Tool: sklearn-crfsuite (https://sklearn-crfsuite.readthedocs.io)

Selected features and gazetteers:

- **F(eatures)**: case and shape features, features for prefixes and suffixes (context window [-2,-1,0,+1,+2])
- **G(azetteers)**: gazetteers of *persons, countries, cities, streets, areas, companies*, *laws, ordinances and administrative regulations* (context window [0])
- **L(ookup)**: lookup table for word similarity with the four most similar words (context window [-2,-1,0,+1,+2])

Developed models:

- CRF-F with features
- CRF-FG with features, gazetteers
- CRF-FGL with features, gazetteers, lookup table

Training parameters:

- Learning algorithm – L-BFGS method
- L1 and L2 regularization parameters (= coefficient 0.1)
- Max. iterations – 100

BiLSTM Training

Tool: UKPLab-BiLSTM (https://github.com/UKPLab/emnlp2017-bilstm-cnn-crf)

Selected models:

- BiLSTM-CRF
- BiLSTM-CRF+ with character embeddings from the BiLSTM
- BiLSTM-CNN-CRF with character embeddings from CNN

Hyperparameters:

- Two BiLSTM (Bidirectional Long-Short Term Memory) layers with a size of 100 units
- Dropout of 0.25
- Max. epochs 100
  pre-trained word embeddings for German

Evaluation

Evaluation:

- Stratified 10-fold cross-validation with shuffling (sentence-wise)
- to prevent overfitting
- to prevent measurement errors in unbalanced data

Measures:

- Micro-Precision
- Micro-Recall
- Micro-F1

Results: CRFs – Fine-grained Classes

Fine-grained classes	CRF-F			CRF-FG			CRF-FGL		
	Prec	**Rec**	**F1**	**Prec**	**Rec**	**F1**	**Prec**	**Rec**	**F1**
Person	89.41	83.53	86.32	90.50	83.54	86.83	90.44	84.22	**87.18**
Judge	98.22	97.62	97.92	98.68	97.75	**98.21**	98.55	97.75	98.14
Lawyer	93.14	76.84	**83.73**	89.81	73.51	80.39	92.17	75.04	81.99
Country	96.73	90.42	93.44	97.03	91.98	94.40	96.93	92.62	**94.70**
City	88.99	77.37	82.70	88.27	81.77	**84.77**	88.09	81.82	84.67
Street	88.69	59.58	70.51	87.51	57.95	68.90	90.50	59.85	**71.30**
Landscape	94.34	61.14	73.43	92.63	64.09	75.25	93.33	65.27	**76.08**
Organization	86.82	71.25	78.20	86.71	71.95	78.56	88.84	72.72	**79.89**
Company	92.77	86.04	89.21	93.00	86.18	89.39	93.54	86.85	**90.01**
Institution	92.74	89.49	**91.07**	92.88	89.20	90.98	92.51	89.47	90.96
Court	97.23	96.35	**96.78**	97.03	96.35	96.69	97.19	96.33	96.75
Brand	85.85	56.91	67.85	90.33	56.20	68.82	88.40	58.07	**69.61**
Law	96.86	96.34	96.60	97.00	96.44	96.72	97.02	96.56	**96.79**
Ordinance	91.91	82.23	86.79	91.35	82.85	86.87	91.41	83.49	**87.26**
European legal norm	89.37	86.07	87.67	88.91	85.49	87.14	**89.41**	86.21	87.76
Regulation	83.83	71.38	77.00	84.34	71.03	**77.02**	84.42	70.66	76.85
Contract	90.66	87.72	**89.15**	90.18	87.42	88.76	90.53	87.67	89.06
Court decision	93.35	93.39	**93.37**	93.22	93.34	93.28	93.21	93.29	93.25
Legal literature	92.98	91.28	92.12	92.94	91.42	**92.17**	92.79	91.28	92.02
Total	94.28	91.85	93.05	94.31	91.96	93.12	94.37	92.12	**93.23**

Results: CRFs – Coarse-grained Classes

Coarse-grained classes	CRF-F			CRF-FG			CRF-FGL		
	Prec	**Rec**	**F1**	**Prec**	**Rec**	**F1**	**Prec**	**Rec**	**F1**
Person	94.20	89.43	91.74	94.54	89.99	**92.20**	94.22	90.20	92.16
Location	94.60	84.55	89.26	93.89	85.48	89.45	94.33	86.45	**90.18**
Organization	92.82	89.00	90.87	93.02	89.08	90.99	93.23	89.10	**91.11**
Legal norm	96.19	95.16	95.67	96.29	95.26	95.77	96.28	95.44	**95.86**
Case-by-case regulation	89.29	84.72	86.94	89.28	84.77	**86.96**	88.76	84.15	86.39
Court decision	93.19	93.26	93.23	93.28	93.23	**93.25**	93.08	93.08	93.08
Legal literature	92.72	91.15	91.92	92.99	91.14	92.06	93.11	91.13	**92.11**
Total	94.17	92.07	93.11	94.26	92.20	**93.22**	94.22	92.25	**93.22**

Results: BiLSTMs – Fine-grained Classes

Fine-grained classes	BiLSTM-CRF			BiLSTM-CRF+			BiLSTM-CNN-CRF		
	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1
Person	89.30	91.08	90.09	90.78	92.24	**91.45**	90.21	92.57	91.35
Judge	98.64	99.48	**99.05**	98.37	99.21	98.78	98.18	99.01	98.59
Lawyer	94.85	84.62	**88.19**	86.18	90.59	87.07	88.02	87.96	87.11
Country	94.66	95.98	95.29	96.52	96.81	**96.66**	95.09	97.20	96.12
City	81.26	86.32	83.48	82.58	89.06	**85.60**	83.21	87.95	85.38
Street	81.70	75.94	78.10	81.82	75.78	77.91	86.24	78.21	**81.49**
Landscape	78.54	79.08	77.57	78.50	80.20	78.25	80.93	81.80	**80.90**
Organization	79.50	74.72	76.89	82.70	80.18	81.28	84.32	81.00	**82.51**
Company	85.81	81.34	83.44	90.05	88.11	89.04	91.72	89.18	**90.39**
Institution	88.88	90.91	89.85	89.99	92.40	91.17	90.24	92.23	**91.20**
Court	97.49	98.33	97.90	97.72	98.24	**97.98**	97.52	98.34	97.92
Brand	78.34	73.11	75.17	83.04	76.25	**79.17**	83.48	73.62	77.79
Law	96.59	97.01	96.80	98.34	98.51	**98.42**	98.44	98.38	98.41
Ordinance	82.63	72.61	77.08	92.29	92.96	**92.58**	91.00	91.09	90.98
European legal norm	90.62	89.79	90.18	92.16	92.63	**92.37**	91.58	92.29	91.92
Regulation	75.58	68.91	71.77	85.14	78.87	**81.63**	79.43	78.30	78.74
Contract	87.12	85.86	86.48	92.00	92.64	**92.31**	90.78	92.06	91.40
Court decision	96.34	96.47	96.41	96.70	96.73	96.71	97.04	97.06	**97.05**
Legal literature	93.87	93.68	93.77	94.34	93.94	94.14	94.25	94.22	**94.23**
Total	93.80	93.70	93.75	95.36	95.57	**95.46**	95.34	95.58	**95.46**

Results: BiLSTMs – Coarse-grained Classes

Coarse-grained classes	BiLSTM-CRF			BiLSTM-CRF+			BiLSTM-CNN-CRF		
	Prec	**Rec**	**F1**	**Prec**	**Rec**	**F1**	**Prec**	**Rec**	**F1**
Person	94.34	95.16	94.74	94.82	96.03	**95.41**	94.09	96.21	95.12
Location	90.85	92.59	91.68	92.60	94.05	**93.31**	91.74	93.45	92.57
Organization	91.82	90.94	91.37	92.87	92.89	92.87	93.80	92.65	**93.21**
Legal norm	97.04	96.50	96.77	97.93	98.04	**97.98**	97.71	97.87	97.79
Case-by-case regulation	86.79	84.15	85.43	90.72	90.53	**90.61**	90.11	90.80	90.43
Court decision	96.54	96.58	96.56	96.93	97.05	**96.99**	96.73	96.83	96.78
Legal literature	93.78	93.91	93.84	94.23	94.62	**94.42**	94.24	93.80	94.02
Total	94.86	94.49	94.68	95.84	96.07	**95.95**	95.71	95.87	95.79

Discussion 1/2

Results per model family

- **BiLSTMs show better performance**

  Good performance for classes that are only covered poorly in the dataset

  BiLSTMs with character embeddings produce **95.46** F1 for fine-grained classes

  BiLSTMs with character embeddings produce **95.95** F1 for coarse-grained classes

- **CRFs are also good**

  CRF with gazetteers and lookup produce **93.23** F1 for fine-grained classes

  CRFs with gazetteers or gazetteers and lookup produce **93.22** F1 for coarse-grained classes

but

  CRFs are about 1-10 F1 lower per class compared to BiLSTMs

  CRFs have bigger differences in precision and recall

Discussion 2/2

Results per class

- For *judge*, *court* and *law* – 95 F1
- For *country*, *institution*, *company*, *court decision*, and *legal literature* – 90 F1
- For *person*, *ordinance*, *European legal norm*, *contract* – 87 F1 (CRFs) and 92 F1 (BiLSTMs)
- For *lawyer* – max. 83.73 F1 (CRFs) and max. 88.19 F1 (BiLSTMs)
- For *city* – max. 85-86 F1
- For *street*, *area*, *organization* and *regulation* – 69–80 F1 (CRFs) and 72–83 F1 (BiLSTMs)
- For *brand* – 69.61 F1 (CRFs) and 79.17 F1 (BiLSTMs)

➢ Good results can be explained by good coverage (*law, court decision*) in the dataset, small number of instances (*judge*, *court*, *country* and *institution*) or uniform citation style (*law*, *court decision*, *legal literature*)

➢ Bad results can be explained by poor coverage (*lawyer*, *street*, *area* and *brand*), heterogeneous representation (*street*, *area*, *organization*), inconsistent citation styles (*regulation and contract)*

Future Work

- Extend or optimize the unbalanced data (to minimize specific influencing factors of data on models)
- Compare these results with new state of the art approaches (language models – preliminary results: no improvements)
- Annotate the dataset by one or two other linguists (planned dataset paper for LREC 2020)

Thank you! Questions?

Project expo includes Lynx booth

Will include an industry workshop (not announced yet)