Using CRF to Detect Different Functional Types of Content in Decisions of US Courts with Example Application to Sentence Boundary Detection

Jaromir Savelka
Kevin D. Ashley

Intelligent Systems Program
University of Pittsburgh

jas438@pitt.edu

June 16, 2017

Even basic NLP tasks such as sentence boundary detection (SBD) are challenging when performed on court decisions.

This is a problem because these techniques are foundational to many applications from IR to computer assisted reasoning.

We assume that one of the root causes of these difficulties is a heterogeneity of content.

- sentences
- lists using punctuation, enumerations, colons
- references and citations
- speech transcripts
- editorial marks

We further assume that the ability to distinguish different types of content may dramatically improve the quality of the processing.

Court Decision's Characteristics

In terms of length a decision may be short (comparable to a newspaper article) or long (similar to a book).

A decision may be structured into sections and subsections preceded by a heading (possibly numbered).

A decision may contain specific constituents such as a header and a footer, footnotes, lists.

Sentences are interleaved with citations.

Sentences themselves may be extremely long, even organized as lists.

There is a high usage of sentence organizers such as ; or — and brackets (multiple types).

Quotes (possibly nested) are frequent.

Presentation Overview

Annotation Scheme

Data Set

Predicting Labels Automatically

Results

Example Application on Sentence Boundary Detection

Annotation Scheme

- Sentence: We have recognized that even a limited search of the person is a substantial invasion of privacy.
- Incomplete sentence: The nature of the writ.
- Non-sentential piece of content: See 387 U.S. 294 (1967).
- External reference: [469 U.S. 325, 329]
- Internal reference: Ante, at 337
- Quotation: "a school official may properly conduct a search of a students person if the official has a reasonable suspicion . . ."
- Heading: ORDER ON A MOTION TO DISMISS
- Numbering token: III; [Footnote 1]
- Editorial Mark: *153; . . .
- Meta data field: Filed: April 16th, 2009

Presentation Overview

Annotation Scheme

Data Set

Predicting Labels Automatically

Results

Example Application on Sentence Boundary Detection

Court Decisions

We downloaded the decisions from the online Court Listener service.

- 13 cyber crime decisions (cyber bullying, credit card frauds, possession of electronic child pornography)
- 3 landmark SCOTUS decisions
- 3 intellectual property protection decisions

	cyber-train	cyber-test	scotus-test	ip-test
# of docs	10	3	3	3
# of tokens	219484	75158	130065	42382
longest doc	181009	121910	190352	54430
average doc	58237	67820	118749	38549
shortest doc	16859	29278	47450	16546

Distribution of Annotations

	SENT	ISENT	NSENT	EREF	IREF	QUOT	EMARK	HEAD	NMB	MDF
cyber-train										
# of seq	3493	495	1945	1574	189	959	818	108	360	43
avg # per doc	349	50	195	157	19	96	82	11	36	4
avg seq length	145	50	31	37	12	143	6	23	3	40
cyber-test										
# of seq	1060	125	728	753	46	325	245	21	109	19
avg # per doc	353	42	243	254	15	108	82	7	36	6
avg seq length	163	84	29	30	7	112	4	33	3	63
scotus-test										
# of seq	1984	158	1316	1256	185	375	164	34	182	2
avg # per doc	661	53	439	419	62	125	55	11	61	1
avg seq length	153	54	34	38	8	123	6	29	3	26
ip-test										
# of seq	608	119	474	506	37	224	116	38	89	2
avg # per doc	203	40	158	169	12	75	39	13	30	1
avg seq length	160	58	26	27	3	111	6	28	3	91
total										
# of seq	7145	897	4463	4099	457	1883	1343	201	740	66
avg # per doc	376	47	235	216	24	99	71	11	39	3
avg seq length	151	57	31	34	9	130	6	26	3	48

Presentation Overview

Sequence Tagging Model

We use conditional random fields (CRF) which is a random field model that is globally conditioned on an observation sequence O.

The states of the model correspond to event labels E.

In our experiments we use a first-order CRF where observation O_i is associated with E_i.

For each annotation type we use a separate model but we use the same training strategy and features for all the models.

We work with the CRFsuite[1] implementation of the first-order CRF.

[1] www.chokkan.org/software/crfsuite/

Training Data and Features

We use the cyber-train data set to train a CRF model for each of the 10 annotation types.

Our tokenization strategy is very aggressive – text is segmented into a greater number of tokens than usual.

Tokenization Example

|Call| |me| |at| |9|am| |on| |my| |phone| |(|123|)|456|-|7890|.|

Each of the tokens is then a data point in a sequence a CRF model operates on.

Training Data and Features II

Each token is represented by a small set of relatively simple features.

1. atfront, atback
2. lower
3. sig
4. length
5. islower, isupper, istitle
6. isdigit, isspace

We also include *lower*, *sig*, *islower*, *isupper*, *istitle*, *isdigit*, *isspace* from the five preceding and five following tokens.

If one of the tokens falls beyond the document boundaries we signal this by including *BOS* and *EOS* features.

Labels

As labels we use the annotation types projected into the BILOU scheme.

Telephone (TEL) and time (TIM) example

|Call| |me| |at| |9|am| |on| |my| |phone| |(|123|)|456|-|7890|.|
O, O, O, O, O, O, B-TIM, L-TIM, O, O, O, O, O, O, O, B-TEL, I-TEL, I-TEL, I-TEL, I-TEL, L-TEL, O

In our work we use one annotation type for each model.

Each of the models was trained on the tag set with 5 labels such as the following one:
B-SENT, I-SENT, L-SENT, O, U-SENT

Presentation Overview

Results

	SENT		ISENT		NSENT		EREF		IREF	
	F_1	supp	F_1	supp	F_1	supp	F_1	supp	F_1	supp
cyber-test										
B	.93	1058	.40	121	.88	645	.85	564	.78	45
I	.96	57630	.42	3400	.94	10483	.95	565	.43	117
L	.93	1058	.44	121	.88	646	.60	565	.80	45
O	.86	15412	.98	71512	.99	63307	.99	10496	1.0	74951
U	–	–	0.0	4	.93	77	0.0	194	–	–
AVG	.94	75158	.95	75158	.98	75158	.98	75158	1.0	75158
scotus-test										
B	.95	1976	.25	144	.80	1165	.74	1107	.44	139
I	.98	99514	.27	2625	.92	20377	.90	19923	.51	540
L	.95	1977	.39	144	.81	1157	.47	1102	.13	139
O	.94	26598	.99	127139	.98	107215	.98	107785	1.0	129201
U	–	–	0.0	13	.70	151	0.0	148	.29	46
AVG	.97	130065	.97	130065	.97	130065	.96	130065	.99	130065
ip-test										
B	.93	600	.38	105	.81	468	.81	445	.91	32
I	.97	31675	.32	2094	.87	6026	.92	6297	.71	50
L	.93	603	.47	105	.86	472	.51	447	.85	37
O	.90	9504	.98	40067	.98	35416	.98	35135	1.0	42263
U	–	–	0.0	11	–	–	.32	58	–	–
AVG	.96	42382	.94	42382	.96	42382	.97	42382	1.0	42382

Results II

	QUOT		EMARK		HEAD		NMB		MDF	
	F_1	supp	F_1	supp	F_1	supp	F_1	supp	F_1	supp
cyber-test										
B	.73	325	.83	98	.62	16	.90	109	.89	19
I	.69	325	.76	142	.69	180	.89	66	.87	370
L	.75	11807	.84	107	.29	16	.90	109	.83	19
O	.94	62701	1.0	74718	1.0	74941	1.0	74874	1.0	74750
U	–	–	.16	93	.75	5	–	–	–	–
AVG	.90	75158	1.0	75158	1.0	75158	1.0	75158	1.0	75158
scotus-test										
B	.49	374	.66	102	.32	23	.50	2	.24	90
I	.41	15496	.58	242	.43	221	.60	14	.30	68
L	.51	373	.69	102	.47	21	.50	2	.24	90
O	.91	113822	1.0	129603	1.0	129789	1.0	130047	1.0	129725
U	–	–	.54	16	.71	11	–	–	.87	92
AVG	.85	130065	1.0	130065	1.0	130065	1.0	130065	1.0	130065
ip-test										
B	.73	224	.70	56	.09	31	.88	89	0.0	2
I	.69	8208	.80	133	.17	290	.84	60	0.0	69
L	.73	224	.71	57	.47	30	.86	88	0.0	2
O	.91	33726	1.0	42113	1.0	42027	1.0	42145	1.0	42382
U	–	–	.82	23	.22	4	–	–	–	–
AVG	.87	42382	1.0	42382	.99	42382	1.0	42382	1.0	42382

Presentation Overview

Sentence Boundary Detection (SBD)

The goal of SBD is to split a natural language text into individual sentences (i.e., identify each sentence's boundaries).

Typically, SBD is operationalized as a binary classification of a fixed number of candidate boundary points (**.**, **!**, **?**).

. . . |Because Mr**.** Lange offered information that was unknown to both competitors and to the general public, there is no reason to confront this issue in this case**.**| |The panel majority bases its reading of the federal statute on its conclusion that . . .

SBD could be a critical task in many applications such as machine translation, summarization, or information retrieval.

Approaches to SBD

- Rules – A battery of hand-crafted matching rules is applied.

IF “!” OR “?” MATCHED $\rightarrow$ MARK AS BOUND
IF “<EOL><EOL>” MATCHED $\rightarrow$ MARK AS BOUND

- Supervised Machine Learning (ML) – Given a triggering event occurs decide if it is an instance of sentence boundary.

$x_i =<$ 0:token=“.”, 0:isTrigger=1, -1:token=“Mr”, -1:isAbbr=1, 1:token=“Lange”, 1:isName=1 $>$
$f(x_i) \rightarrow y_i$

- Unsupervised ML – Similar to supervised ML approach but the system is trained on unlabeled data.

SBD Performance

Multiple SBD systems were reported as having an excellent performance:[1]

- 99.8% accuracy of a tree-based classifier in predicting "." as ending (or not) a sentence evaluated on Brown corpus[Riley 1989]
- 99.5% accuracy of a combination of original system based on neural nets and decision trees with existing system[Aberdeen&al. 1995] evaluated on WSJ corpus[Palmer&Hearst 1997]
- 99.75% accuracy (WSJ) and 99.64% (Brown) of a maximum entropy model in assessing ".", "!", and "?"[Reynar&Ratnaparkhi 1997]
- 99.69% (WSJ) and 99.8% (Brown) of a rule-based sentence splitter combined with a supervised POS-tagger[Mikheev 2002]
- 98.35% (WSJ) and 98.98% (Brown) of an unsupervised system based on identification of abbreviations[Kiss&Strunk 2006]

The list comes from Read et al. 2012.

SBD Performance (cont.)

	Type	Brown	CDC	Genia	WSJ	All
CoreNLP	R	87.7 (93.6)	72.1 (98.3)	98.8 (99.0)	91.3 (94.8)	89.1 (95.0)
LingPipe$_1$	R	94.9 (96.6)	87.6 (99.1)	98.3 (98.6)	97.3 (98.7)	95.2 (97.4)
LingPipe$_2$	R	93.0 (94.5)	86.3 (97.2)	**99.6** (**99.8**)	88.0 (90.9)	93.2 (95.3)
MxTerminator	S	94.7 (96.5)	97.9 (98.6)	98.3 (98.5)	97.4 (98.5)	95.8 (97.2)
OpenNLP	S	96.6 (96.6)	98.6 (98.6)	98.8 (98.8)	99.1 (99.1)	**97.4** (97.4)
Punkt	U	96.4 (96.4)	**98.7** (98.7)	99.3 (99.3)	98.3 (98.3)	97.3 (97.3)
RASP	R	**96.8** (96.8)	96.1 (99.1)	98.9 (98.9)	99.0 (99.0)	**97.4** (**97.6**)
Splitta	S	95.4 (95.4)	96.1 (96.7)	99.0 (99.0)	**99.2** (**99.2**)	96.5 (96.5)
tokenizer	R	94.9 (**96.9**)	98.6 (**99.2**)	98.6 (98.9)	97.9 (**99.2**)	96.2 (**97.6**)

	WeScience		WNB		WLB	
	A	B	A	B	A	B
CoreNLP	90.0	97.9	95.3	96.4	89.1	90.9
LingPipe$_1$	90.0	98.1	94.8	96.1	92.4	94.2
LingPipe$_2$	89.8	98.0	94.4	95.6	92.7	94.5
MxTerminator	89.5	97.2	94.7	95.9	90.3	92.2
OpenNLP	90.2	97.9	95.3	96.5	90.2	92.0
Punkt	89.9	97.7	**95.6**	96.7	92.8	94.5
RASP	**91.0**	99.1	95.4	96.6	92.8	94.6
Splitta	**91.0**	98.9	94.0	95.5	91.2	93.4
tokenizer	**91.0**	**99.2**	**95.6**	**96.8**	**93.1**	**94.9**

Experiments from Read et al. 2012

Court Decision Excerpt

A long sentence with a quote (with a nested quote) organized as a list followed by citations and a short sentence with a citation.

. . . As used in the statute, "'act in furtherance of a person's right of petition or free speech under the United States or California Constitution in connection with a public issue' includes: (1) any written or oral statement or writing made before a legislative, executive, or judicial proceeding, or any other official proceeding authorized by law; (2) any written or oral statement or writing made in connection with an issue under consideration or review by a legislative, executive, or judicial body, or any other official proceeding authorized by law; (3) any written or oral statement or writing made in a place open to the public or a public forum in connection with an issue of public interest; (4) or any other conduct in furtherance of the exercise of the constitutional right of petition or the constitutional right of free speech in connection with a public issue or an issue of public interest." (§425.16, subd. (e), italics added; see Briggs v. Eden Council for Hope & Opportunity (1999) 19 Cal. 4th 1106, 1117-1118, 1123 [81 Cal.Rptr.2d 471, 969 P.2d 564] [discussing types of statements covered by anti-SLAPP statute].) The R.'s contend that plaintiffs' complaint falls within the third clause of section 425.16, subdivision (e). . . .

Court Decision Excerpt

Semicolons separate items in a list as well as independent clauses.

[O]ur family suffered: emotional distress; anxiety; sleeplessness; physical pain; insecurity; fear; pain and suffering; payment of attorneys' fees; payment of medical expenses; payment of moving expenses; payment of *1204 traveling and housing expenses to and from Los Angeles to support our business endeavors; [and] [D.C.]'s lost income. . . .

Completed assemblies must be exhaustively tested to demonstrate, to the FAA's satisfaction, that all requirements have been met; only then does the FAA certify the part for sale.

It takes RAPCO a year or two to design, and obtain approval for, a complex part; the dynamometer testing alone can cost $75,000. . . . Drawings and other manufacturing information contain warnings of RAPCO's intellectual-property rights; every employee receives a notice that the information with which he works is confidential.

Court Decision Excerpt

Informal poorly edited text may be present.

The next post, from "DAN JUSTICE," is the first to raise the rhetoric to a level that could, when considered out of context, be construed as a threat. It says "HEY [D.C.], I KNOW A GOOD *** WHEN I SEE ONE. I LIKE WHAT I SEE, LET'S GO GET SOME COFFEE. ****** im gonna kill you" and is signed "H-W student."

A sentence may span over a double line break.

. . . Section 1(4) of the Uniform Act provides:

"Trade secret" means information, including a formula, pattern, compilation, program, device, method, technique, or process . . .

Heading (possibly no triggering event)

FACTS AND PROCEDURAL HISTORY

SBD Data Set

We use the cyber-train, cyber-test, scotus-test, and ip-test data sets for the SBD experiments.

We do not use the concept of triggering events; any character can potentially be a boundary.

We apply a consistent policy of aggressive segmenting (i.e., if doubts exist there is a boundary).

	cyber-train	cyber-test	scotus-test	ip-test
# of docs	10	3	3	3
# of sentences	5423	1757	3214	1107
longest sentence	1065	1182	670	1145
average sentence	106	114	110	103
shortest sentence	1	1	1	2

Evaluated SBD Systems

For evaluation of SBD systems' performance on the corpus of court decisions we use one system from each category:

1. We work with the SBD module from the Stanford CoreNLP toolkit[Manning&al. 2014] as an example of a system based on rules.[1]
2. To test a system based on supervised ML classifier we employ the SBD component from openNLP.[2]
3. As an example of an unsupervised system we use the punkt[Kiss&Strunk 2006] module from the NLTK toolkit.[4]

The criterion for selection of the SBD systems was the assumed wide adoption of general toolkits the SBD systems are part of.

[1] `nlp.stanford.edu/software/corenlp.shtml`
[2] `opennlp.apache.org`
[3] `nltk.org/api/nltk.tokenize.html`

Evaluation

We use traditional IR measures – precision (P), recall (R), and F_1-measure (F_1).

We evaluate the SBD performance from two different perspectives:

1. boundaries
2. segments – both boundaries need to match

For each perspective we use two approaches to determine if the boundary was predicted correctly.

1. strict – boundary offsets match exactly
2. lenient – the difference between boundary offsets does not contain alphanumeric char

|Accordingly, we find that the circuit court did not abuse its discretion when it denied Mr.| |Renfrow's motion for a JNOV.| |**|We find no merit to this issue.|

Performance of Off-the-shelf SBD Systems

	cyber-test			scotus-test			ip-test		
	P	R	F1	P	R	F1	P	R	F1
CoreNLP									
strict-B	.81	.78	.79	.87	.76	.82	.75	.79	.77
lenient-B	.82	.78	.80	.88	.77	.82	.75	.79	.77
strict-S	.56	.54	.55	.70	.60	.64	.50	.53	.52
lenient-S	.57	.55	.56	.70	.61	.65	.51	.54	.52
openNLP									
strict-B	.88	.77	.82	.84	.74	.78	.79	.78	.79
lenient-B	.88	.77	.82	.84	.74	.78	.80	.79	.79
strict-S	.64	.56	.60	.65	.57	.61	.57	.55	.56
lenient-S	.64	.56	.60	.66	.58	.61	.57	.56	.56
punkt									
strict-B	.72	.79	.75	.77	.72	.74	.67	.80	.73
lenient-B	.72	.79	.75	.78	.73	.76	.69	.83	.75
strict-S	.41	.46	.44	.55	.52	.54	.42	.50	.46
lenient-S	.42	.47	.44	.56	.53	.55	.44	.53	.48

Error Analysis

Missed boundary following a unit if a triggering event is absent

B. Response to Jury Question|

Deliberate avoidance is not a standard less than knowledge;| it is simply another way that knowledge may be proven.

Missed boundaries between citations

Kolender v. Lawson, 461 U.S. 352, 357, 103 S. Ct. 1855, 75 L. Ed. 2d 903 (1983);| United States v. Lim, 444 F.3d 910, 915 (7th Cir.2006)

Wrongly predicted boundaries in citations

see United States v. X-Citement Video, Inc., 513 U.S. 64, 76-78, 115 S. Ct. 464, 130 L. Ed.| 2d 372 (1994)

Performance of Trained SBD Systems

	cyber-test			scotus-test			ip-test		
	P	R	F1	P	R	F1	P	R	F1
CoreNLP++									
strict-B	.81	.90	.85	.86	.93	.89	.75	.88	.81
lenient-B	.81	.91	.86	.86	.94	.90	.75	.88	.81
strict-S	.62	.70	.66	.71	.77	.74	.52	.62	.57
lenient-S	.63	.70	.66	.72	.78	.75	.53	.62	.57
openNLP++									
strict-B	.93	.79	.85	.91	.76	.83	.94	.81	.87
lenient-B	.93	.79	.85	.91	.76	.83	.94	.82	.87
strict-S	.71	.59	.64	.74	.61	.67	.72	.62	.67
lenient-S	.71	.60	.65	.74	.62	.68	.73	.62	.67
punkt++									
strict-B	.77	.79	.78	.82	.71	.77	.76	.80	.78
lenient-B	.77	.79	.78	.84	.73	.78	.79	.83	.81
strict-S	.47	.49	.48	.61	.53	.57	.52	.55	.53
lenient-S	.47	.49	.48	.62	.54	.58	.54	.57	.55

Custom CRF Model

We train a custom CRF model which is essentially the same as the models described earlier.

In addition to the 10 features we add the automatically predicted labels corresponding to the following types:

1. Sentence
2. Incomplete Sentence
3. Non-sentential Sequence

This means we use a two-phased SBD system based on CRF.

1. the three CRF models attempt to label the sequence of tokens in terms of the three above mentioned types.
2. the single CRF model uses the information from the first pass to predict the sentence boundaries.

We use cyber-train as the training set.

Performance of Trained SBD Systems

	cyber-test			scotus-test			ip-test		
	P	R	F1	P	R	F1	P	R	F1
CoreNLP++									
strict-B	.81	.90	.85	.86	.93	.89	.75	.88	.81
lenient-B	.81	.91	.86	.86	.94	.90	.75	.88	.81
strict-S	.62	.70	.66	.71	.77	.74	.52	.62	.57
lenient-S	.63	.70	.66	.72	.78	.75	.53	.62	.57
openNLP++									
strict-B	.93	.79	.85	.91	.76	.83	.94	.81	.87
lenient-B	.93	.79	.85	.91	.76	.83	.94	.82	.87
strict-S	.71	.59	.64	.74	.61	.67	.72	.62	.67
lenient-S	.71	.60	.65	.74	.62	.68	.73	.62	.67
Custom CRF									
strict-B	.94	.96	.95	.90	.95	.92	.95	.94	.95
lenient-B	.94	.96	.95	.91	.95	.93	.96	.95	.95
strict-S	.86	.86	.86	.81	.85	.83	.86	.84	.85
lenient-S	.86	.87	.86	.82	.85	.83	.86	.84	.85

Error Analysis

The judgment of the Court of Appeals for the D. C. Circuit is affirmed.

. . . a "case we have described as a 'monument of English freedom' 'undoubtedly familiar' to 'every American statesman' at the time the Constitution was adopted . . .

. . . search is not involved and resort must be had to Katz analysis; but there is no reason for rushing forward . . .

Conclusions

We tested the hypothesis that one of the reasons why court decisions are challenging for NLP processing is the heterogeneity of content.

We classified the content of selected decisions in terms of 10 different types with varying degree of success.

On an example application of SBD we have shown how the information about the content type improves the performance of an SBD system.

Thank you!

Questions, comments and suggestions are welcome now or any time at jas438@pitt.edu.

References I

Aberdeen, J., Burger, J., Day, D., Hirschman, L., Robinson, P., & Vilain, M. (1995, November). MITRE: description of the Alembic system used for MUC-6. In *Proceedings of the 6th conference on Message understanding* (pp. 141–155). Association for Computational Linguistics.

de Maat, Emile, and Radboud Winkels. "A next step towards automated modelling of sources of law." *Proceedings of the 12th International Conference on Artificial Intelligence and Law*. ACM, 2009.

de Maat, Emile, Radboud Winkels, and Tom Van Engers. "Automated detection of reference structures in law." *Frontiers in Artificial Intelligence and Applications* (2006): 41.

Dell'Orletta, Felice, et al. "The SPLeT-2012 shared task on dependency parsing of legal texts." *Proceedings of the 4th Workshop on Semantic Processing of Legal Texts*. 2012.

Grabmair, M., Ashley, K. D., Chen, R., Sureshkumar, P., Wang, C., Nyberg, E., and Walker, V. R. (2015, June). Introducing LUIMA: an experiment in legal conceptual retrieval of vaccine injury decisions using a UIMA type system and tools. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law* (pp. 69-78). ACM.

References II

- Kiss, Tibor, and Jan Strunk. "Unsupervised multilingual sentence boundary detection." Computational Linguistics 32.4 (2006): 485-525.
- John Lafferty, Andrew McCallum, and Fernando Pereira. “Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data". Proceedings of the 18th International Conference on Machine Learning. 282–289. 2001.
- Liu, Y., Stolcke, A., Shriberg, E., & Harper, M. (2005, June). Using conditional random fields for sentence boundary detection in speech. In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (pp. 451-458). Association for Computational Linguistics.
- Manning, Christopher D., Surdeanu, Mihai, Bauer, John, Finkel, Jenny, Bethard, Steven J., and McClosky, David. 2014. The Stanford CoreNLP Natural Language Processing Toolkit In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60.
- Mikheev, Andrei. “Periods, capitalized words, etc.” Computational Linguistics 28.3 (2002): 289-318.
- Okazaki, Naoaki. “CRFsuite: a fast implementation of conditional random fields (CRFs).” (2007).

References III

- Palmer, David D., and Marti A. Hearst. "Adaptive multilingual sentence boundary disambiguation." Computational Linguistics 23.2 (1997): 241-267.
- Ratnaparkhi, Adwait. Maximum entropy models for natural language ambiguity resolution. Diss. University of Pennsylvania, 1998.
- Read, J., Dridan, R., Oepen, S., & Solberg, L. J. (2012). Sentence boundary detection: A long solved problem?. COLING (Posters), 12, 985-994.
- Reynar, Jeffrey C., and Adwait Ratnaparkhi. "A maximum entropy approach to identifying sentence boundaries." Proceedings of the fifth conference on Applied natural language processing. Association for Computational Linguistics, 1997.
- Riley, Michael D. 1989. Some applications of tree-based modelling to speech and language. In Proceedings of the workshop on Speech and Natural Language (HLT '89). Association for Computational Linguistics, Stroudsburg, PA, USA, 339-352.
- Tran, Oanh Thi, Minh Le Nguyen, and Akira Shimazu. "Reference resolution in legal texts." *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law*. ACM, 2013.
- Tran, O. T., Ngo, B. X., Le Nguyen, M., and Shimazu, A. (2014). Automated reference resolution in legal texts. *Artificial intelligence and law*, 22(1), 29-60.

References IV

- Opijnen, Marc van, Nico Verwer, and Jan Meijer. "Beyond the Experiment: the eXtendable Legal Link eXtractor." *ASAIL* (2015).
- Wyner, Adam, and Wim Peters. "On Rule Extraction from Regulations." *JURIX*. Vol. 11. 2011.
- *CRFsuite: A fast implementation of Conditional Random Fields (CRFs)* [online]. Accessed on 2017-03-14 at `http://www.chokkan.org/software/crfsuite/`
- *GATE: general architecture for text engineering* [online]. Accessed on 2017-03-14 at `http://gate.ac.uk`
- *LingPipe* [online]. Accessed on 2017-03-14 at `http://alias-i.com/lingpipe/`
- *MxTerminator* [online]. Accessed on 2017-03-14 at `ftp://ftp.cis.upenn.edu/pub/adwait/jmx/`
- *openNLP* [online]. Accessed on 2017-03-14 at `http://opennlp.apache.org/`
- *NLTK 3.0*: punkt module [online]. Accessed on 2017-03-14 at `http://nltk.org/api/nltk.tokenize.html`
- *iLexIR*: RASP [online]. Accessed on 2017-03-14 at `https://www.ilexir.co.uk/nlp-applications/index.html`

References V

splitta [online]. Accessed on 2017-03-14 at `https://code.google.com/archive/p/splitta/`

Stanford CoreNLP – a suite of core NLP tools [online]. Accessed on 2017-03-14 at `http://nlp.stanford.edu/software/corenlp.shtml`

tokenizer [online]. Not accessible on 2017-03-14 at `http://www.cis.uni-muenchen.de/~wastl/misc/`