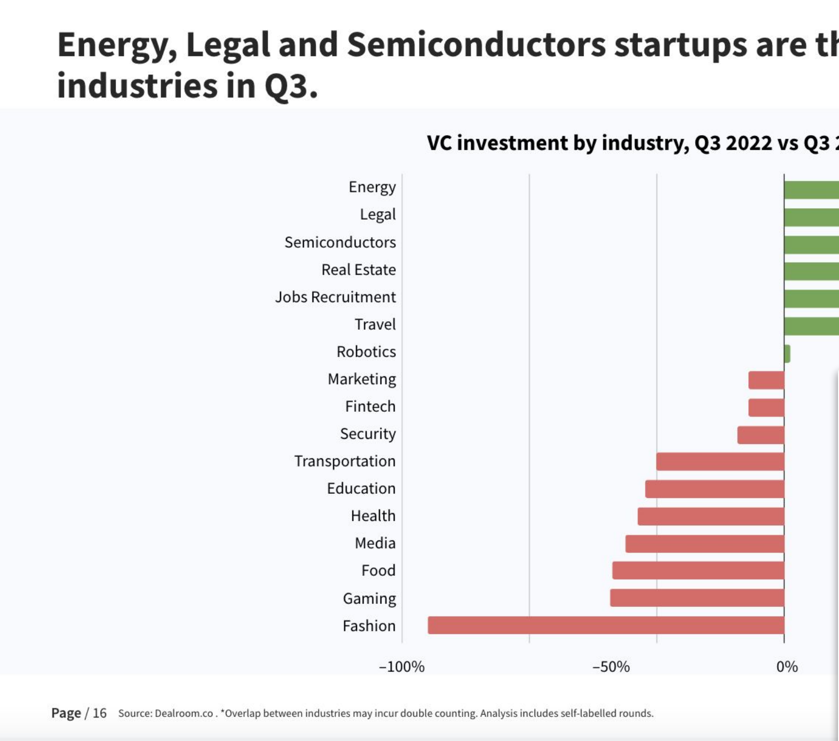
Legal NLP applications

György Orosz

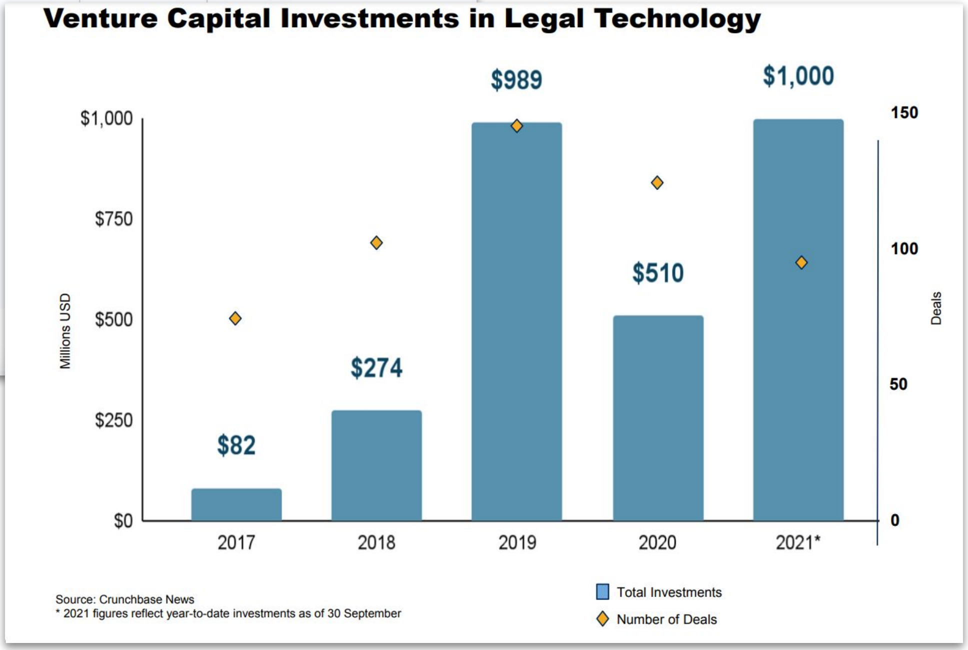
gyorgy@orosz.link



Energy, Legal and Semiconductors startups are the fastest-growing tech industries in Q3.

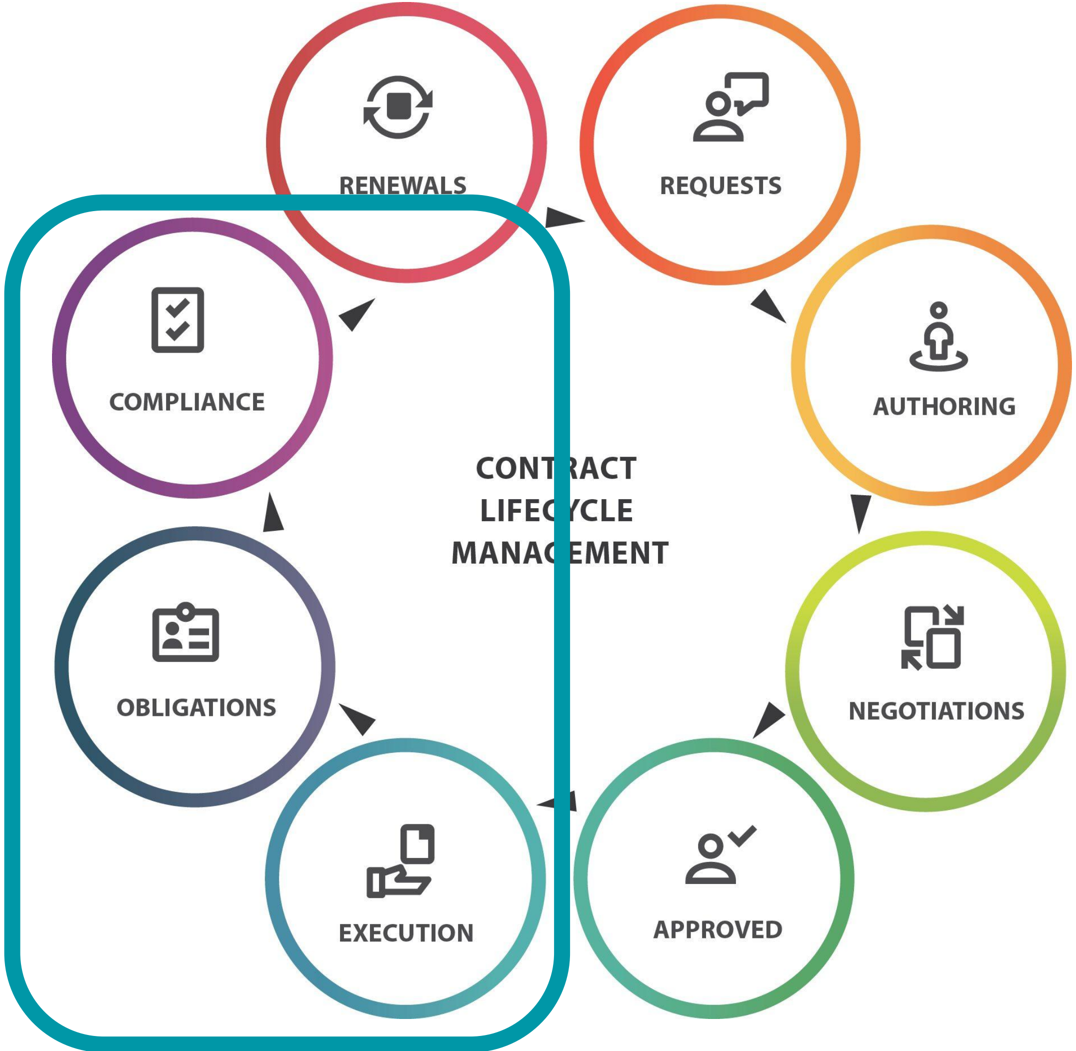
VC investment by industry, Q3 2022 vs Q3 2021

Page / 16 Source: Dealroom.co . *Overlap between industries may incur double counting. Analysis includes self-labelled rounds.



Venture Capital Investments in Legal Technology

Source: Crunchbase News
* 2021 figures reflect year-to-date investments as of 30 September


RENEWALS
REQUESTS
COMPLIANCE
AUTHORING
CONTRACT
LIFECYCLE
MANAGEMENT
OBLIGATIONS
NEGOTIATIONS
EXECUTION
APPROVED

Contracts @ Execution / Management

- Document management
- Contract review
- Compliance
- Tracking obligations
- Due diligence
- Legal research
- Electronic discovery

Kira

Affinitext
MAKING DOCUMENTS INTELLIGENT

eBrevia
by DFIN

ThoughtRiver™
contract acceleration

ROSS

Luminance

P

casetext

definely

Avvoka

superlegal

DEMO

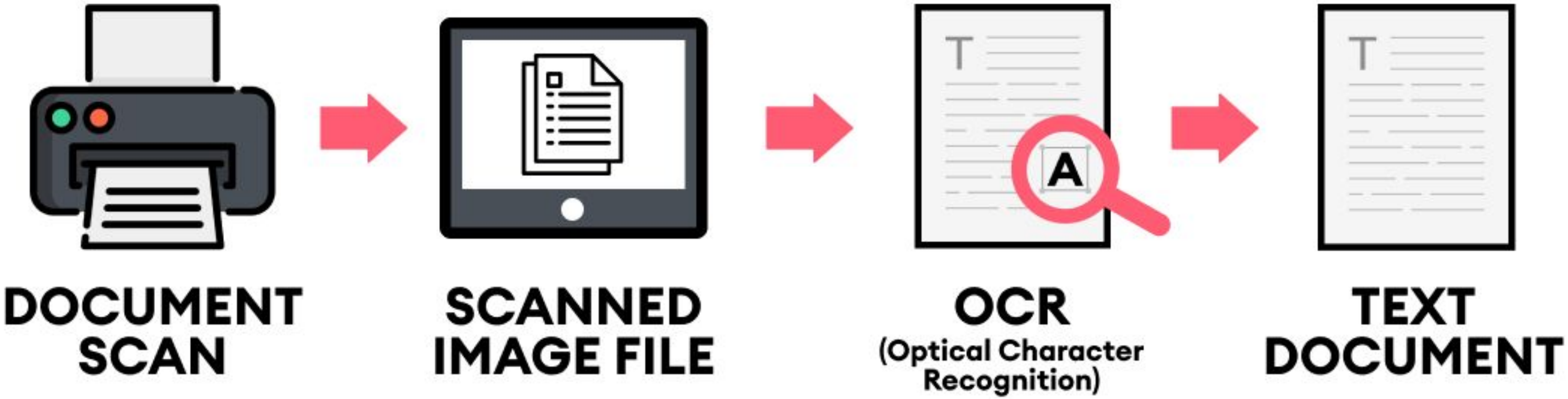
Legal language

- Tend to have more structure (e.g. hierarchical numbering)
- Is more precise (lawyers are rewarded for reducing ambiguity)
- Has a smaller / specific vocabulary
- Grammatically well constructed
- There is strong domain knowledge

Legal documents...

- are long
- build on explicit definitions which are often specified elsewhere in the document
- extensively use citations to other documents or document parts

Is not a formal language but a natural one: despite many attempts to bring formal logic to the aid of legal writing, the law remains a domain of natural language semantics.



DOCUMENT SCAN

SCANNED IMAGE FILE

OCR (Optical Character Recognition)

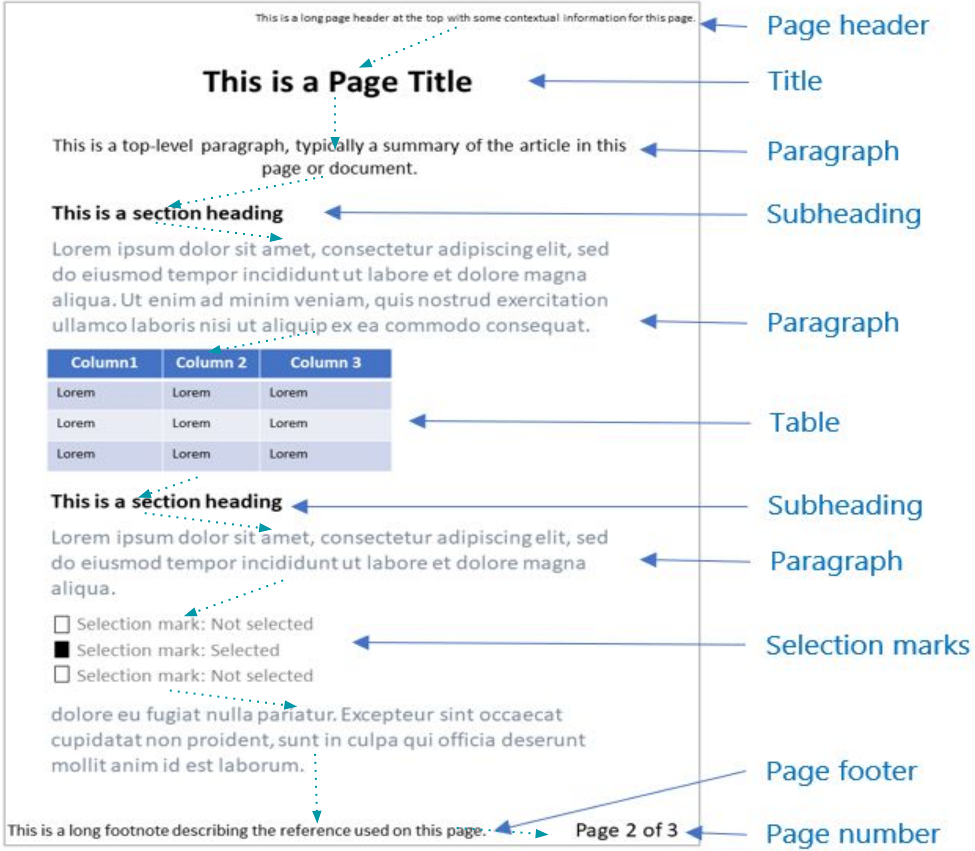
TEXT DOCUMENT

PDF

Adobe

Document layout analysis

Huang, Yupan, et al. "Layoutlmv3: Pre-training for document ai with unified text and image masking." Proceedings of the 30th ACM International Conference on Multimedia. 2022.



This is a long page header at the top with some contextual information for this page.

This is a Page Title

This is a top-level paragraph, typically a summary of the article in this page or document.

This is a section heading

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

Column1	Column 2	Column 3
Lorem	Lorem	Lorem
Lorem	Lorem	Lorem
Lorem	Lorem	Lorem

This is a section heading

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

□ Selection mark: Not selected
■ Selection mark: Selected
□ Selection mark: Not selected

dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

This is a long footnote describing the reference used on this page.

Page 2 of 3

https://huggingface.co/spaces/deepdoctection/deepdoctection

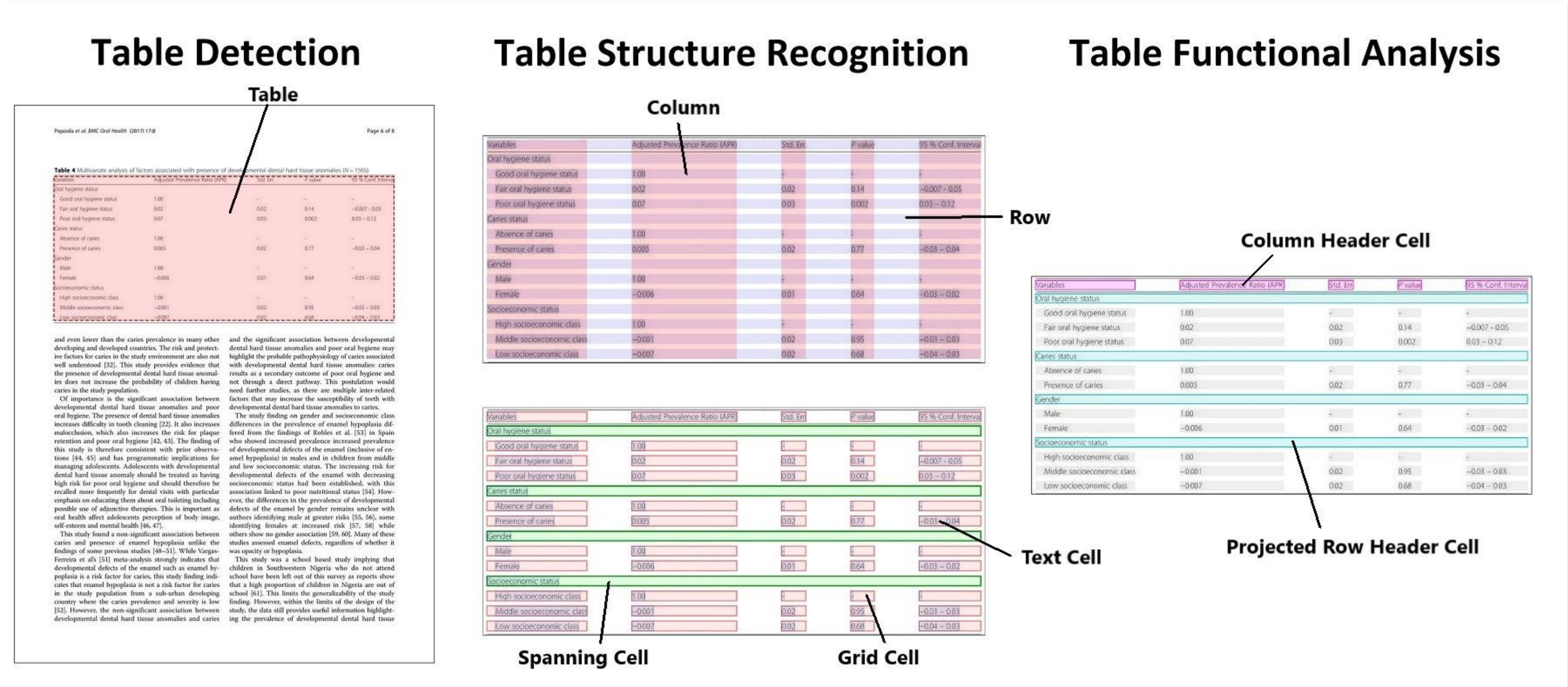
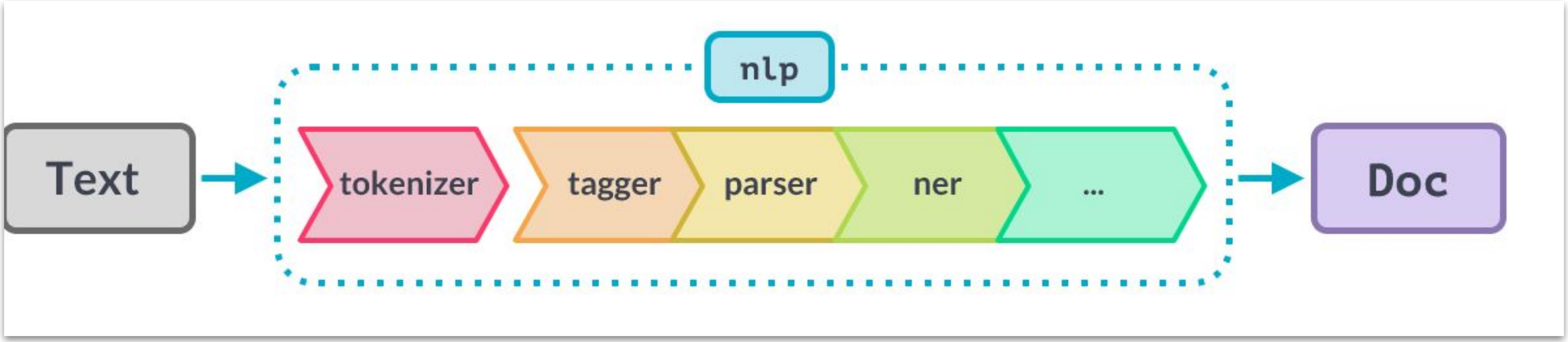


Table 4 Multivariate analysis of factors associated with presence of developmental dental hard tissue anomalies (N = 1565)

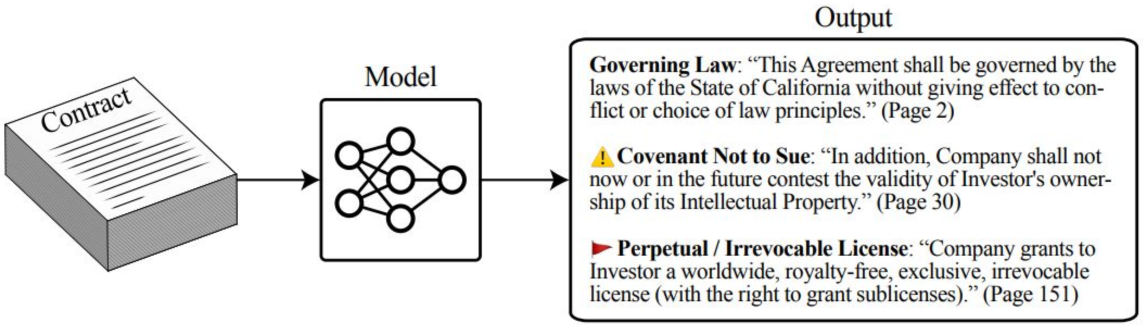
Variables	Adjusted Prevalence Ratio (APR)	Std. Err.	P value	95 % Conf. Interval
Oral hygiene status				
Good oral hygiene status	1.00	-	-	-
Fair oral hygiene status	0.02	0.02	0.14	−0.007 - 0.05
Poor oral hygiene status	0.07	0.03	0.002	0.03 – 0.12
Caries status				
Absence of caries	1.00	-	-	-
Presence of caries	0.005	0.02	0.77	−0.03 – 0.04
Gender				
Male	1.00	-	-	-
Female	−0.006	0.01	0.64	−0.03 – 0.02
Socioeconomic status				
High socioeconomic class	1.00	-	-	-
Middle socioeconomic class	−0.001	0.02	0.95	−0.03 – 0.03
Low socioeconomic class	−0.007	0.02	0.68	−0.04 – 0.03

Smock, Brandon, Rohith Pesala, and Robin Abraham. "PubTables-1M: Towards comprehensive table extraction from unstructured documents." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

Preprocessing texts



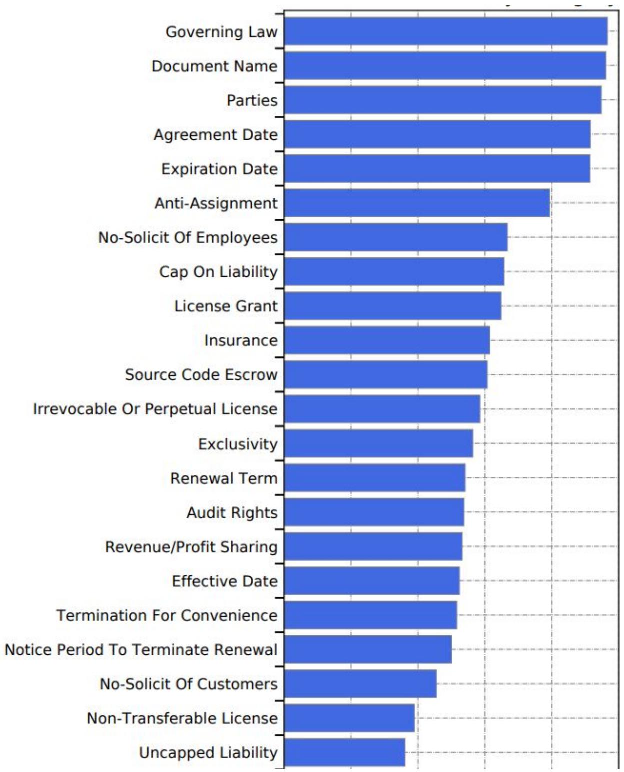
Legal element extraction



Governing Law: “This Agreement shall be governed by the laws of the State of California without giving effect to conflict or choice of law principles.” (Page 2)

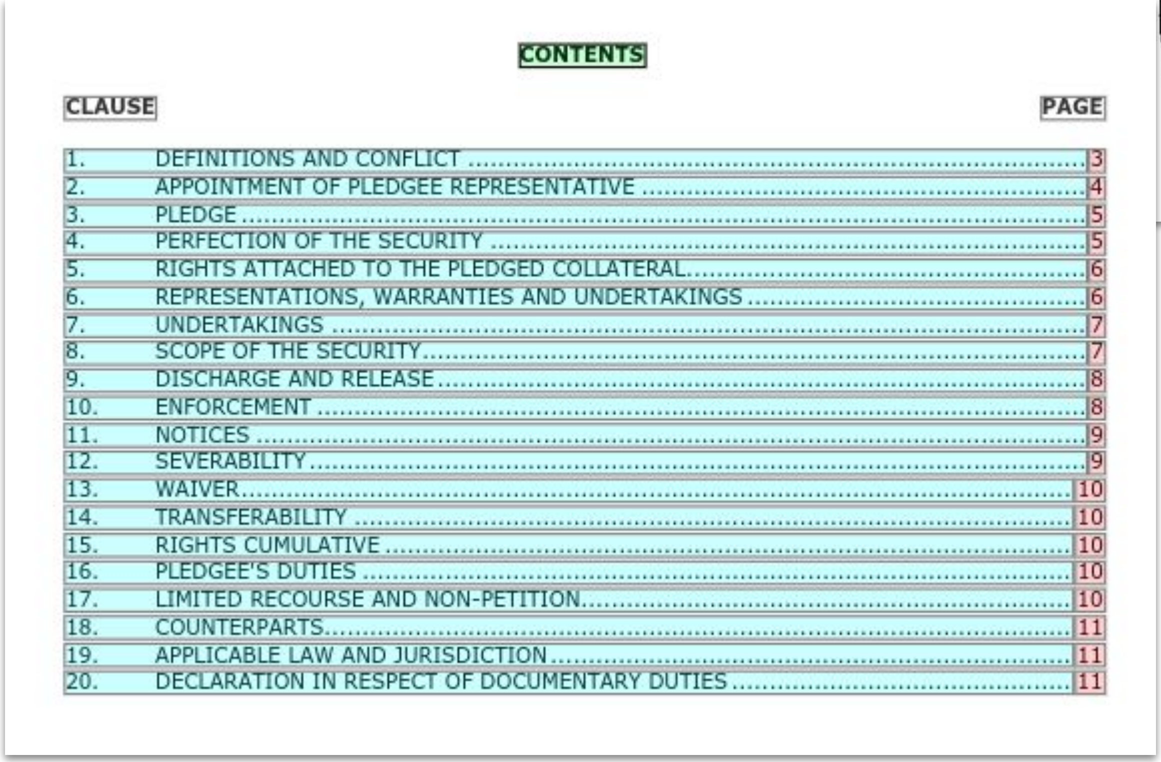
⚠ **Covenant Not to Sue**: “In addition, Company shall not now or in the future contest the validity of Investor's ownership of its Intellectual Property.” (Page 30)

► **Perpetual / Irrevocable License**: “Company grants to Investor a worldwide, royalty-free, exclusive, irrevocable license (with the right to grant sublicenses).” (Page 151)



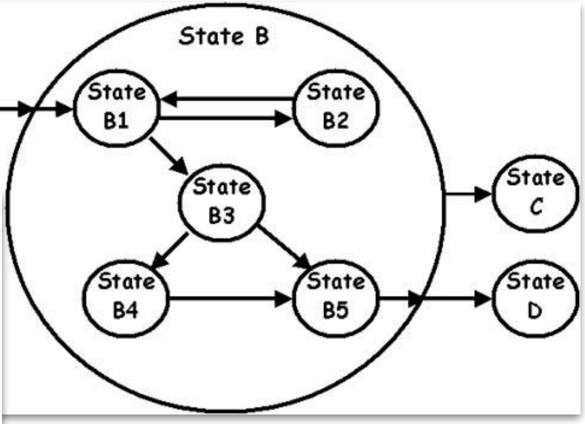
Hendrycks, Dan, et al. "Cuad: An expert-annotated nlp dataset for legal contract review." arXiv preprint arXiv:2103.06268 (2021).

Table of contents detection

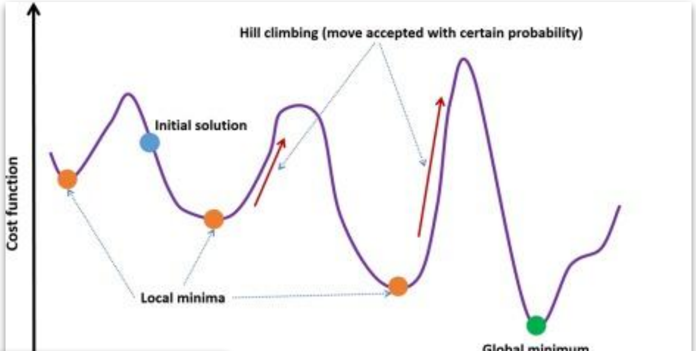


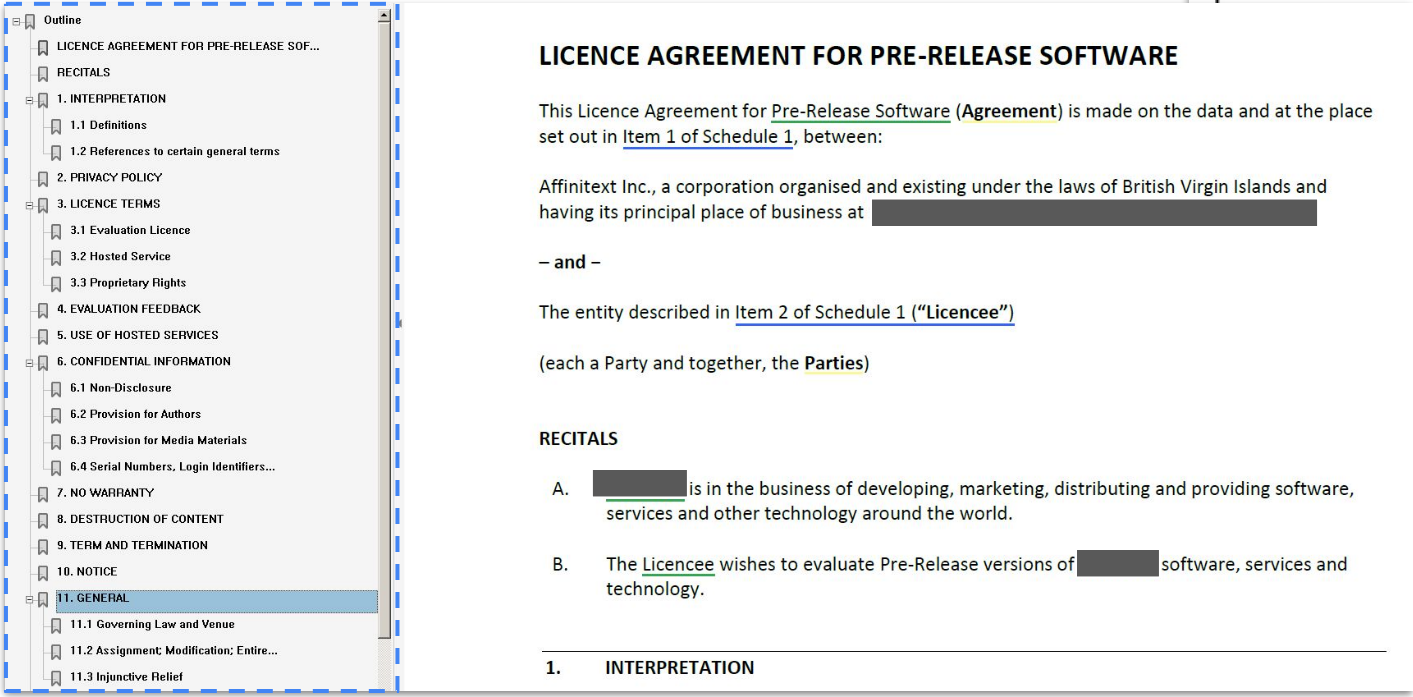
CONTENTS

CLAUSE PAGE



Document structure detection





- Outline
  - LICENCE AGREEMENT FOR PRE-RELEASE SOF...
  - RECITALS
  - 1. INTERPRETATION
    - 1.1 Definitions
    - 1.2 References to certain general terms
  - 2. PRIVACY POLICY
  - 3. LICENCE TERMS
    - 3.1 Evaluation Licence
    - 3.2 Hosted Service
    - 3.3 Proprietary Rights
  - 4. EVALUATION FEEDBACK
  - 5. USE OF HOSTED SERVICES
  - 6. CONFIDENTIAL INFORMATION
    - 6.1 Non-Disclosure
    - 6.2 Provision for Authors
    - 6.3 Provision for Media Materials
    - 6.4 Serial Numbers, Login Identifiers...
  - 7. NO WARRANTY
  - 8. DESTRUCTION OF CONTENT
  - 9. TERM AND TERMINATION
  - 10. NOTICE
  - 11. GENERAL
    - 11.1 Governing Law and Venue
    - 11.2 Assignment; Modification; Entire...
    - 11.3 Injunctive Relief

LICENCE AGREEMENT FOR PRE-RELEASE SOFTWARE

This Licence Agreement for Pre-Release Software (**Agreement**) is made on the data and at the place set out in Item 1 of Schedule 1, between:

Affinitext Inc., a corporation organised and existing under the laws of British Virgin Islands and having its principal place of business at [redacted]

– and –

The entity described in Item 2 of Schedule 1 (**"Licencee"**)

(each a Party and together, the **Parties**)

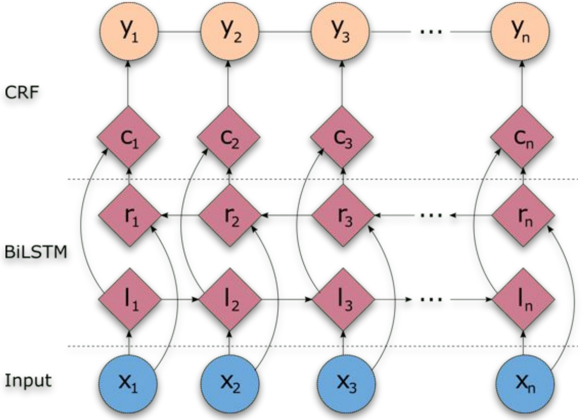
RECITALS

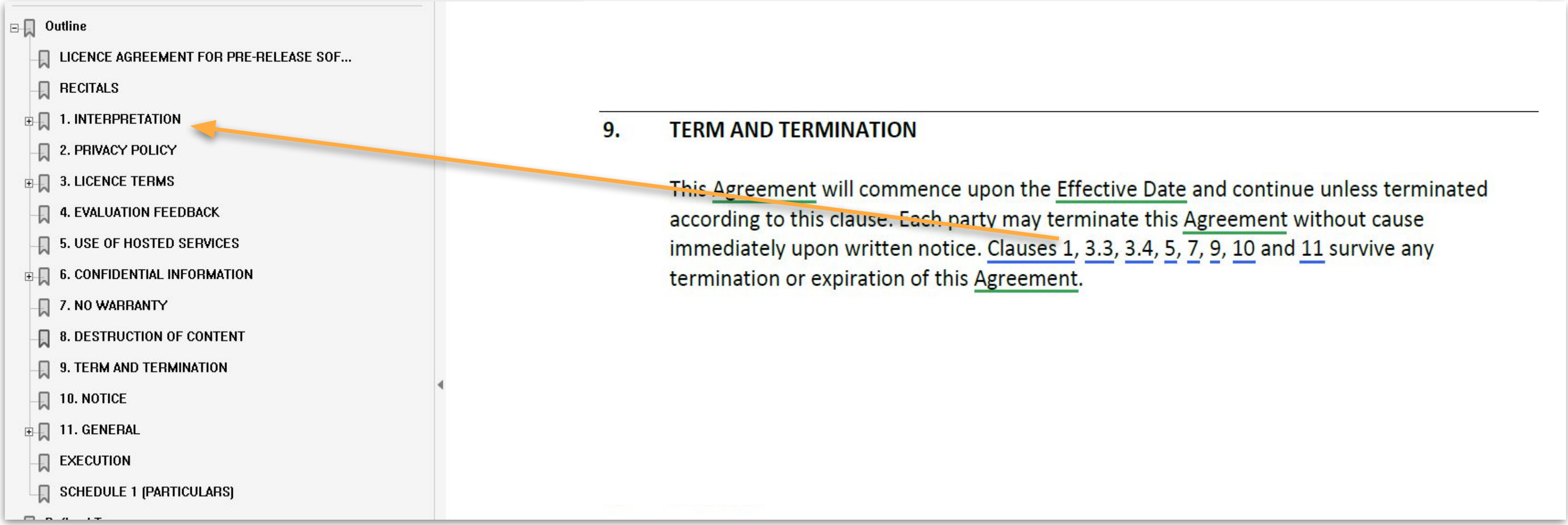
A. [redacted] is in the business of developing, marketing, distributing and providing software, services and other technology around the world.

B. The Licencee wishes to evaluate Pre-Release versions of [redacted] software, services and technology.

1. INTERPRETATION

Cross-references





Outline

LICENCE AGREEMENT FOR PRE-RELEASE SOF...

RECITALS

1. INTERPRETATION

2. PRIVACY POLICY

3. LICENCE TERMS

4. EVALUATION FEEDBACK

5. USE OF HOSTED SERVICES

6. CONFIDENTIAL INFORMATION

7. NO WARRANTY

8. DESTRUCTION OF CONTENT

9. TERM AND TERMINATION

10. NOTICE

11. GENERAL

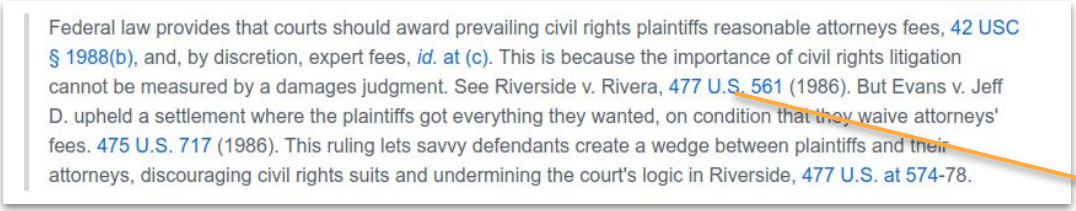
EXECUTION

SCHEDULE 1 (PARTICULARS)

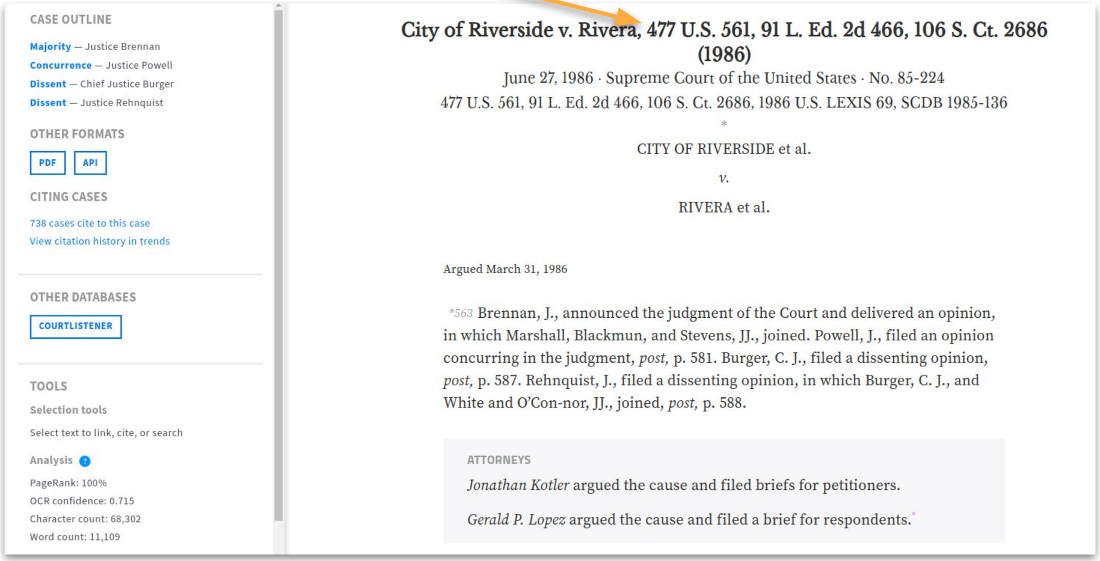
9. TERM AND TERMINATION

This Agreement will commence upon the Effective Date and continue unless terminated according to this clause. Each party may terminate this Agreement without cause immediately upon written notice. Clauses 1, 3.3, 3.4, 5, 7, 9, 10 and 11 survive any termination or expiration of this Agreement.

Legal citations



> Federal law provides that courts should award prevailing civil rights plaintiffs reasonable attorneys fees, 42 USC § 1988(b), and, by discretion, expert fees, *id.* at (c). This is because the importance of civil rights litigation cannot be measured by a damages judgment. See Riverside v. Rivera, 477 U.S. 561 (1986). But Evans v. Jeff D. upheld a settlement where the plaintiffs got everything they wanted, on condition that they waive attorneys' fees. 475 U.S. 717 (1986). This ruling lets savvy defendants create a wedge between plaintiffs and their attorneys, discouraging civil rights suits and undermining the court's logic in Riverside, 477 U.S. at 574-78.



CASE OUTLINE

Majority — Justice Brennan
Concurrence — Justice Powell
Dissent — Chief Justice Burger
Dissent — Justice Rehnquist

OTHER FORMATS

CITING CASES

738 cases cite to this case
View citation history in trends

OTHER DATABASES

TOOLS

Selection tools

Select text to link, cite, or search

Analysis

PageRank: 100%
OCR confidence: 0.715
Character count: 68,302
Word count: 11,109

City of Riverside v. Rivera, 477 U.S. 561, 91 L. Ed. 2d 466, 106 S. Ct. 2686 (1986)

June 27, 1986 · Supreme Court of the United States · No. 85-224
477 U.S. 561, 91 L. Ed. 2d 466, 106 S. Ct. 2686, 1986 U.S. LEXIS 69, SCDB 1985-136

*

CITY OF RIVERSIDE et al.

v.

RIVERA et al.

Argued March 31, 1986

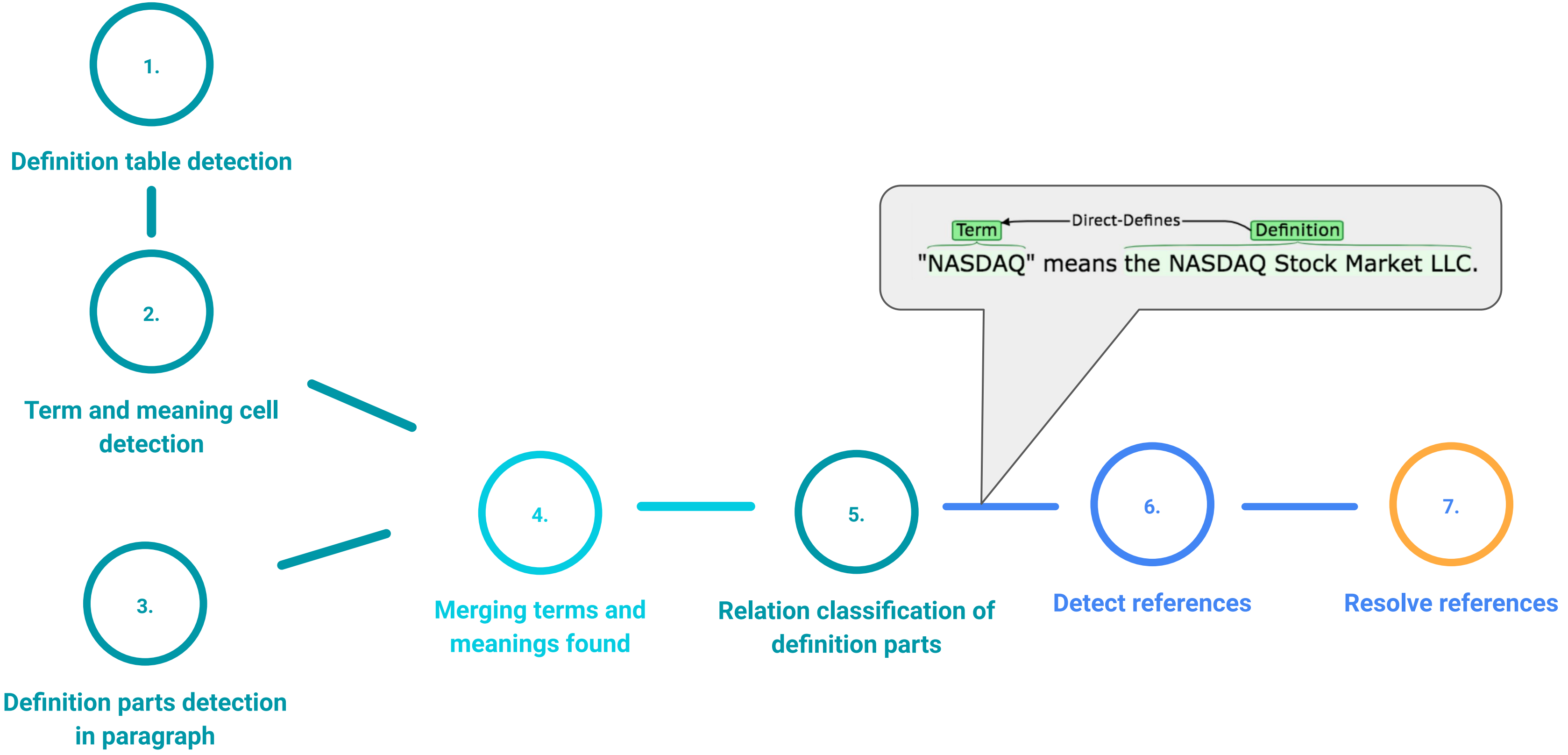
*563 Brennan, J., announced the judgment of the Court and delivered an opinion, in which Marshall, Blackmun, and Stevens, JJ., joined. Powell, J., filed an opinion concurring in the judgment, *post,* p. 581. Burger, C. J., filed a dissenting opinion, *post,* p. 587. Rehnquist, J., filed a dissenting opinion, in which Burger, C. J., and White and O'Con-nor, JJ., joined, *post,* p. 588.

ATTORNEYS

Jonathan Kotler argued the cause and filed briefs for petitioners.

Gerald P. Lopez argued the cause and filed a brief for respondents.*

Definition related tasks



1. Definition table detection

2. Term and meaning cell detection

3. Definition parts detection in paragraph

4. Merging terms and meanings found

5. Relation classification of definition parts

"NASDAQ" means the NASDAQ Stock Market LLC.

6. Detect references

7. Resolve references

DeftEval competition results

Sasha Spala, Nicholas Miller, Franck Dernoncourt, and Carl Dockhorn. 2020. SemEval-2020 Task 6: Definition Extraction from Free Text with the DEFT Corpus. In Proceedings of the Fourteenth Workshop on Semantic Evaluation, pages 336–345, Barcelona (online). International Committee for Computational Linguistics.

- Term & meaning identification: 84.71%
- Relation classification: 99.43%

Term detection on our own dataset:

- LexNLP: 44.89%
- Ours: 92.75%

Thank you!