Ravel Data Challenge

We'd like you to think about how you would design a system to automatically extract **relationships between entities** mentioned in natural language. We've given you ten judicial opinions that involve certain companies, represented by certain law firms. For each company, your system should find:

- The set of **law firms** that represent the company (if any)
- The company's **party type(s)** in the litigation (as *plaintiff, defendant, …*)

The input to your system already includes tagged mentions of companies and law firms, but not of party types. For example, Opinion #3 includes this line:

> *For {partyA_}ZENITH ELECT CORPORATION{/partyA_}, a Delaware corporation, plaintiff: John M. Calimafde, {firm1_}Hopgood, Calinafde, Kalil, Blaustein & Judlowe{/firm1_}, New York, NY. Joel S. Feldman, {firm2_}Sachnoff & Weaver, Ltd.{/firm2_}, Chicago, IL.*

On the following page, we've also given you a table with the list of relevant companies in each opinion, and the gold-standard annotated outputs for **party type(s)** and **law firm(s)**. Your system should fill in the two *modeled* columns, trying to match the gold standard results as often as possible. Here's what the table looks like for the complete Opinion #3, including the fact that Party A is a plaintiff represented by Law Firms 1 and 2:

Op	Party	Party name	Party type(s) - Annotated	Law firm(s) - Annotated	Party type(s) - Modeled	Law firm(s) - Modeled
3	A	ZENITH ELECTRONICS CORPORATION	plaintiff, counter-defendant	1,2		
3	B	EXZEC INC.	defendant, counter-claimant, counter-plaintiff	3,4		
3	C	CARROLL TOUCH INC	defendant, counter-claimant, counter-plaintiff	5		

- Not all of the information given in the opinions must be used. For this task, we're only interested in the companies listed in the table, and the law firms that represent them.
- For this exercise, we don't expect a perfectly tuned system, but we are interested in your thoughts about handling the various situations encountered throughout the sample.
- More generally, we'll be interested in your overall thoughts on algorithms, annotations, processes, feasibility, modeling, evaluation, or any other aspect of automating this task.
- Plan to spend 10-15 minutes at your onsite interview sharing your conclusions with a mixed group of technical and non-technical team members.
- Please reach out to mark@ravellaw.com with any questions about this data challenge!

Op	Party Letter	Party name	Party type(s) - Annotated	Law firm(s) - Annotated	Party type(s) - Modeled	Law firm(s) - Model
1	B	National Railroad Passenger Corp.	defendant	2		
1	G	Consolidated Rail Corp.	defendant	2		
2	A	625 3RD ST. ASSOCS., L.P.	plaintiff	1,2		
2	B	ALLIANT CREDIT UNION	defendant	3,4,5		
3	A	ZENITH ELECTRONICS CORPORATION	plaintiff, counter-defendant	1,2		
3	B	EXZEC INC.	defendant, counter-claimant, counter-plaintiff	3,4		
3	C	CARROLL TOUCH INC	defendant, counter-claimant, counter-plaintiff	5		
4	D	Adamowicz Estates	plaintiff	1		
5	A	ALLSTATE INSURANCE COMPANY	plaintiff	1		
6	A	SAMSUNG ELECTRONICS CO., LTD.	appellee	1,2		
6	B	QIMONDA AG	appellee	1,2		
6	C	Infineon Technologies AG	appellee	1,2		
6	D	International Business Machines Corp.	appellee	1,2		
6	E	Hynix Semiconductor, Inc.	appellee	3		
6	F	Intel Corporation	appellee	4		
6	G	Nanya Technology Corporation	appellee	5		
6	H	Micron Technology, Inc.	appellee	6		
7	A	HOME BUILDERS ASSOCIATION OF NORTHERN CALIFORNIA	plaintiff	1		
7	B	UNITED STATES FISH AND WILDLIFE SERVICE	defendant	2		
7	C	California Building Industry Association	plaintiff	1		
7	D	Building Industry Legal Defense Foundation	plaintiff	1		
7	F	United States Department of the Interior	defendant	2		
7	H	Center for Biological Diversity	defendant-intervenor			
8	B	VIEJAS GROUP BARON LONG CAPITAN GRANDE BAND	defendant	2		
9	A	Sharp Electronics Corp.	petitioner	1		
9	B	Branded Products, Inc.	respondent	2,3		
10	A	Nvidia Corporation	defendant	6,7		
10	B	ATI Technologies, Inc.	defendant	6,7		
10	C	Advanced Micro Devices, Inc.	defendant	6,7		
10	D	AMD U.S. Finance, Inc.	defendant	6,7		
10	E	Alberta ULC	defendant	6,7		