# **Fine-Tuned Relationship Extraction for Consumer Goods Concepts**

# Data Science Institute COLUMBIA UNIVERSITY

## **Background:**

The foundation of Machine Learning models is data, and annotating text data for training NLP models is always time consuming. Existing solutions are usually too complex to use. For example, the annotation tool shown in Fig. 1 has lots of buttons, and users need to manually input the sentences, switch back and forth between several tabs to make a single annotation.

Our project aims to create an open-source annotation software which identifies the entities, recognizes the entity pairs and their corresponding relations using previous models to boost the annotation process. In consequence, annotators could focus on uncertainty of current predictions , reducing cost both timely and financially.

	Annotate your	<b>Relational T</b>	riplets					
A	Save 🗅 Loa 🕁	d						
	Sentence	Relation	Add Annotation	Output	Demo	Samples		
	Sentence Inpu	ıt						
								Read
	List of sentend	ces					List of entities	
	Ser	ntence						

Figure 1. Existing Annotation Tool(<u>https://abera87.github.io/annotate/</u>)

# Methods:

Separating frontend and backend is conductive to developing user-interactive functions and making backend focusing on dealing with data when we use models from **Open NRE and spaCy. We choose flask, a lightweight framework which supports single** page applications and provides full control for users, as our backend building tool. For frontend framework, we choose React due to a steep learning curve. From the perspective of performance, React uses virtual DOM and only loads the different parts in the DOM tree after interacting with users, leading a fast loading speed. Fig. 2 demonstrates our wireflow design for the front end.

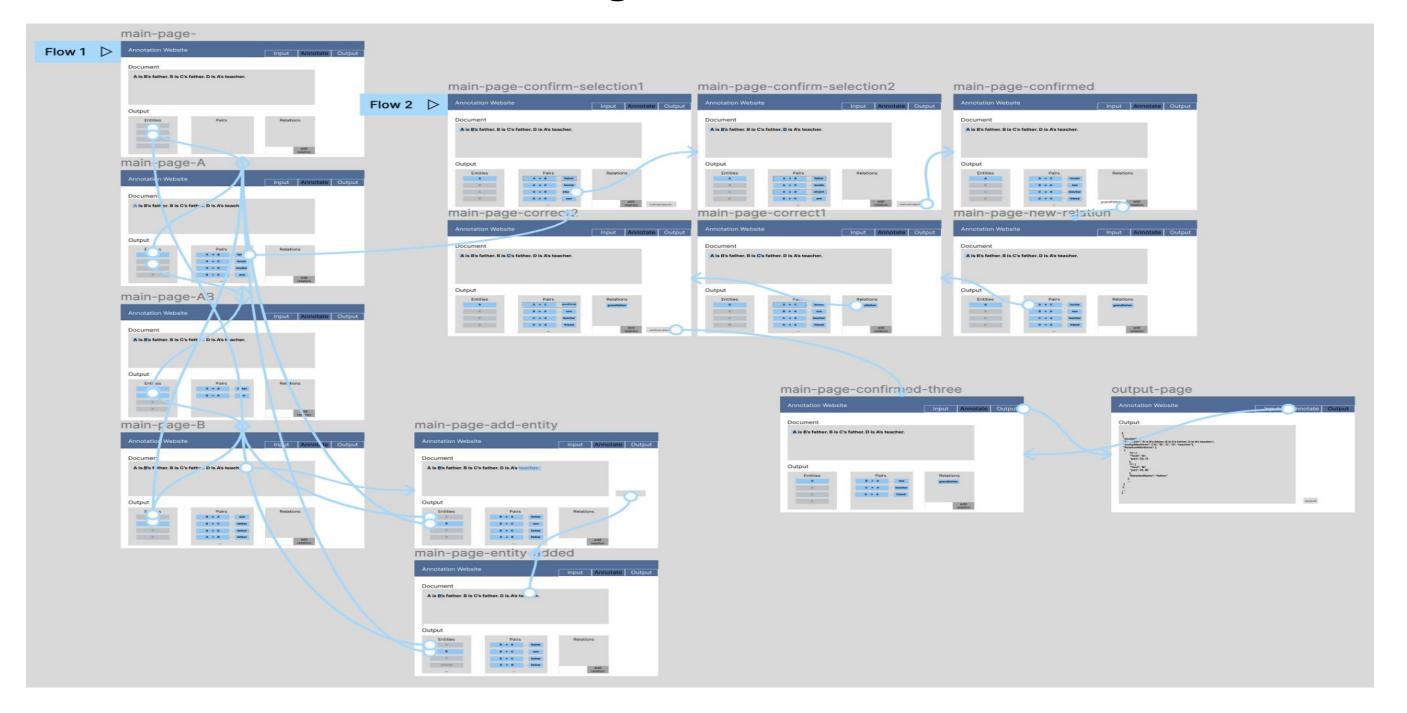


Figure 2. Our wireflow design for frontend

# Wenxin Zhang, Xingyu Lu, Yinghao Li, Jiazhen Li, Wei Luo, **Mentor: John Labarga**

### Results

As is shown in Fig. 3, our tool has three main pages: 1) input page that enables users to input documents through mongoDB connection 2) index page that allows users to select documents to annotate and 3) annotation page that unifies all functions like adding entities, relations, assigning relations and etc.

# Conclusion

Up to now, we have finished implementing the tool and received positive feedback from our monitor. However, whether we need to further improve our website will be depended on the feedbacks from future users. We hope our annotation tool can help boost the process of annotating text contents and make positive impact for the development of NLP field.

### Acknowledgments

This project is supported by Unilever. We would like to express our sincere thank to all those who are related to this project. Primarily, we are grateful to our mentor John Labarga and professor Sining Chen for their support and guidance in completing this project. References

- 1. <u>https://reactjs.org/</u>
- 2. <u>https://vuejs.org/</u>
- 3. <u>https://angular.io/</u>
- November 29, 2022, from
- Retrieved November 29, 2022, from

# **Data Science Capstone Project** with Unilever

		[rest] [mex] [Arrow									
	Denistration I.										
	Current Input							1793a11329ucab93x868	02311793a11379ccab016860		5a/1379ccali/016
Contraction String								en internet fa a manufacturation menung antimisti transfer and an Company Property and photometers	The advancement of feasible charge regime of feasible charge events of	and peak the second sec	ny pomoto ni i lugar a ladin ny disease 2014, angelari
Detailable Harris								Appropriate	Annahalar		Annala
puterned											
Cirlection Name											
abattactis							633	11795af1379ecab61bilbib	633/1296a/1321ccab@tbille	633/179	Gal1379coab61bi
st Pasie.							(BLICTY)	Processing contractions for physics between the second sec	BOXEDID Down index ensure CICI is a signed as perities risking of an annual of a	in and I ferrar card on the	en al colora parateri
Descamed Form									American		Arrende
text								Advitate			
Inpu	it from MangaDB						633	1796an379ccab6/04be	63311796a11379ccah61b6bf	633(179	6ar1379ecab61b6
Consider Strep								anna ridoar fo anni d'agaiste ka	The distance of the plane public companies with solar the The transferio and public of the transferior to the transferior of t	Non-America N	nering 2011 harvede same
								Amintain	domina		Annuals.
Database Norm:											
Colection Name-											
0.0000000000000000000000000000000000000							633	rt796art379ccab61b8c1	433/1798a/1375ccab@lb8c2	632/179	04/137941484184
In Face							-1140	in spine divise of Citity in a succession appr 19 - Titl - solar back of piloting (en. p.	Encoded to the set of the set is for all others and the set of the	10. m	and conserve the second
							prove and a second	Invident			
Dominist Pald									daminin		Amonda
	- 9.4HE						423	1796aft370rcabil/biloit	433/1796a/1379ccab61b8c5	633/178	Parts79ccab4164
ut page foi							Ir			_	
		Annotation Web	site			2	Input Index	Annotate			
	Document	Annotation Web	site			Đ		Annotate			
	Document The shortening of hospi	al stays implies rethinking the pre-	and post-operative mail	Nogement of lower limb arthrop	iesty. Optimal preparatio	n of the patient and antic	Input Index	ative process are necessary	r to limit the length of		
	The shortening of hospil stay and ensure qual		and post-operative mai	a what is known about preoperati r those at risk of complicated por	ise information, educatio	n of the patient and antici n and nehabilitation for pr	Input Index	stive process are necessary	t recommended in		
	The shortening of hospil stay and ensure qual	el stays inglies rethinking the pre- y, safety and patient astisfaction. I	and post-operative mai	what is known about preoperate those at risk of complicates those shortened	ise information, educatio stoperative outcomes. In	n of the patient and antici n and nehabilitation for pr	Input Index	stive process are necessary	t recommended in		
Entities	The shortening of hospil stay and ensure qual	el stays inglies rethinking the pre- y, safety and patient astisfaction. I	and post-operative mai	what is known about preopential those at risk of complicated po- shortened	se information, education topperative outcomes. In liengths of stay.	n of the patient and antici n and nehabilitation for pr	Input Index	elive process are necessary Physical netabilitation is no rate the whole process affe	t recommended in		
Entities	The shortening of hospi size and ensure qual isolation. However, patie	el stays implies rethinking the pre- y, safety and patient astisfaction. I t-centred information and educatio	and post-operative ma This article summarises on is recommended for Pai	what is known about preopential those at risk of complicated po- shortened	se information, educatio topperative outcomes, in liengths of stay.	n of the patient and antic n and rehabilitation for pr tertfisciplinary collaborati	Ingut Index	elive process are necessary Physical netabilitation is no rate the whole process affe	t recommended in		
Entities	The shortening of hospil stay and ensure qual	el stays inglies rethinking the pre- y, safety and patient astisfaction. I	and post-operative ma This article summarises on is recommended for Pai	what is known about preopential those at risk of complicated po- shortened	se information, educatio topperative outcomes, in liengths of stay.	n of the patient and antici n and nehabilitation for pr	Ingut Index	elive process are necessary Physical netabilitation is no rate the whole process affe	t recommended in		
Entities	The shortening of hospitistay and ensure qual- isolation. However, patie	el stays implies rethinking the pre- y, safety and patient astisfaction. I t-centred information and educatio	and post-operative ma This article summarises on is recommended for Pai	what is known about preopential those at risk of complicated po- shortened	se information, educatio topperative outcomes, in liengths of stay.	n of the patient and antic n and rehabilitation for pr tertfisciplinary collaborati	Ingut Index	elive process are necessary Physical netabilitation is no rate the whole process affe	t recommended in		
Entities	The shortening of hospi size and ensure qual isolation. However, patie	el stays implies rethinking the pre- y, safety and patient astisfaction. I t-centred information and educatio	and post-operative ma This article summarises on is recommended for Pai	what is known about preopential those at risk of complicated po- shortened	se information, educatio topperative outcomes, in liengths of stay.	n of the patient and antic n and rehabilitation for pr tertfisciplinary collaborati	Ingut Index	elive process are necessary Physical netabilitation is no rate the whole process affe	t recommended in		
Entities	The shortening of hospitistay and ensure qual- isolation. However, patie	el stays implies rethinking the pre- y, safety and patient astisfaction. I t-centred information and educatio	and post-operative ma This article summarises on is recommended for Pai	what is known about preopential those at risk of complicated po- shortened	se information, educatio topperative outcomes, in liengths of stay.	n of the patient and antic n and rehabilitation for pr tertfisciplinary collaborati	Ingut Index	elive process are necessary Physical netabilitation is no rate the whole process affe	t recommended in		
Entities	The shortening of hospitistay and ensure qual- isolation. However, patie	el stays implies rethinking the pre- y, safety and patient astisfaction. I t-centred information and educatio	and post-operative ma This article summarises on is recommended for Pai	what is known about preopential those at risk of complicated po- shortened	se information, educatio topperative outcomes, in liengths of stay.	n of the patient and antic n and rehabilitation for pr tertfisciplinary collaborati	Ingut Index	elive process are necessary Physical netabilitation is no rate the whole process affe	t recommended in		
Entities	The shortening of hospitistay and ensure qual- isolation. However, patie	el stays implies rethinking the pre- y, safety and patient astisfaction. I t-centred information and educatio	and post-operative ma This article summarises on is recommended for Pai	what is known about preopential those at risk of complicated po- shortened	se information, educatio topperative outcomes, in liengths of stay.	n of the patient and antic n and rehabilitation for pr tertfisciplinary collaborati	Ingut Index	elive process are necessary Physical netabilitation is no rate the whole process affe	t recommended in		
Entities	The shortening of hospitistay and ensure qual- isolation. However, patie	el stays implies rethinking the pre- y, safety and patient astisfaction. I t-centred information and educatio	and post-operative ma This article summarises on is recommended for Pai	what is known about preopential those at risk of complicated po- shortened	se information, educatio topperative outcomes, in liengths of stay.	n of the patient and antic n and rehabilitation for pr tertfisciplinary collaborati	Ingut Index	elive process are necessary Physical netabilitation is no rate the whole process affe	t recommended in		
Entities	The shortening of hospitistay and ensure qual- isolation. However, patie	el stays implies rethinking the pre- y, safety and patient astisfaction. I t-centred information and educatio	and post-operative ma This article summarises on is recommended for Pai	what is known about preopential those at risk of complicated po- shortened	se information, educatio topperative outcomes, in liengths of stay.	n of the patient and antic n and rehabilitation for pr tertfisciplinary collaborati	Ingut Index	elive process are necessary Physical netabilitation is no rate the whole process affe	t recommended in		
Entities	The shortening of hospitistay and ensure qual- isolation. However, patie	el stays implies rethinking the pre- y, safety and patient astisfaction. I t-centred information and educatio	and post-operative ma This article summarises on is recommended for Pai	what is known about preopential those at risk of complicated po- shortened	se information, educatio topperative outcomes, in liengths of stay.	n of the patient and antic n and rehabilitation for pr tertfisciplinary collaborati	Ingut Index	elive process are necessary Physical netabilitation is no rate the whole process affe	t recommended in		
Entities	The shortening of hospitistay and ensure qual- isolation. However, patie	el stays implies rethinking the pre- y, safety and patient astisfaction. I t-centred information and educatio	and post-operative ma This article summarises on is recommended for Pai	what is known about preopential those at risk of complicated po- shortened	se information, educatio topperative outcomes, in liengths of stay.	n of the patient and antic n and rehabilitation for pr tertfisciplinary collaborati	Ingut Index	elive process are necessary Physical netabilitation is no rate the whole process affe	t recommended in		
Entities	The shortening of hospitistay and ensure qual- isolation. However, patie	el stays implies rethinking the pre- y, safety and patient astisfaction. I t-centred information and educatio	and post-operative ma This article summarises on is recommended for Pai	what is known about preopential those at risk of complicated po- shortened	se information, educatio topperative outcomes, in liengths of stay.	n of the patient and antic n and rehabilitation for pr tertfisciplinary collaborati	Ingut Index	elive process are necessary Physical netabilitation is no rate the whole process affe	t recommended in		
Entities	The shortening of hospitistay and ensure qual- isolation. However, patie	el stays implies rethinking the pre- y, safety and patient astisfaction. I t-centred information and educatio	and post-operative ma This article summarises on is recommended for Pai	what is known about preopential those at risk of complicated po- shortened	se information, educatio topperative outcomes, in liengths of stay.	n of the patient and antic n and rehabilitation for pr tertfisciplinary collaborati	Ingut Index	stive process are necessary Physical netwolk process effe	t recommended in		

Annotation page that unifies all function

Figure 3. Main functions of the annotation tool

4. Flask vs Django: Which python framework to choose? InterviewBit. (2022, July 29). Retrieved https://www.interviewbit.com/blog/flask-vs-django/#:~:text=Django%20is%20a%20full%2Dstac

k,Working%20of%20Framework%2FData%20Model

5. Martin, S. (2022, February 3). Angular vs react vs Vue.js: Which is the best choice for 2022. is.

https://plainenglish.io/blog/angular-vs-react-vs-vue-js-which-is-the-best-choice-for-2022

