



# Building Mental Health Knowledge Graph

Xiaoying Zhang, Yunrui Shao

**USC**Viterbi

School of Engineering

# Introduction & Motivation



Need help?



Authority?  
Therapist?  
Medicine?  
Emotional support?

Patients with issue of  
mental health

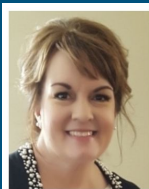


That's our Knowledge Graph!


# Datasets

Total: 13470 pages

## Psychology Today



**Ursula Towles**

Counselor, LCPC, NCC  Verified

Life's storms come to each of us. We can feel overwhelmed, weary, anxious, alone, or stuck. Change, pain, & loss occur or events happen that cause deep trauma. Sometimes even the mundane day-to-day aspects of life can take a toll or addictions rule what was once our lif...

Office is near:  
California, MD 20619

(301) 703-9245

View

crawling



name	person_id	title	about	website	...
Ursula Towles	0	Counselor	"Life's storms ..."	http://www.disc...	...

9956 entries x 14 props

## WIKIDATA Query

Wikidata Query Service

Examples
Query Builder
Help
More tools

```

1 SELECT DISTINCT ?item ?itemLabel ?itemDescription WHERE {
2   {SELECT ?item ?drugs WHERE {
3     ?item wdt:P31|wdt:P279* wd:Q12135.}}
4   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
5 ORDER BY (?itemLabel)

```

cleaning



disease_wiki	disease_name	description	drugs	symptoms	risk_factors
Q55787200	11p15.4 mic	human disease			
Q55788229	11q22.2q22.2	human disease			
Q55788876	12q14 micro	gene: mutations in this gene have already been impl			

2467 entries x 6 props

## Drugs.com

### Amphetamine

Generic name: amphetamine [ *am-FET-a-meen* ]

Brand names: **Adzenys ER**, **Adzenys XR-ODT**, **Dyanavel XR**, **Evekeo**, **Evekeo ODT**, Amphetamine Sulfate

Dosage forms: oral suspension, extended release (1.25 mg/mL; 2.5 mg/mL); oral tablet (10 mg; 5 mg); oral tablet, disintegrating (10 mg; 15 mg; 20 mg; 5 mg); oral tablet, disintegrating, extended release (12.5 mg; 15.7 mg; 18.8 mg; 3.1 mg; 6.3 mg; 9.4 mg)

Drug class: **CNS stimulants**

crawling



drug_id	drug_name	description	warnings	drug_classes	brands
https://www	Amphetamir	Amphetamir	Amphetamir	CNS stimula	Adzenys ER,
https://www	Zopiclone	Zopiclone is	Zopiclone may cause seric	Imovane, Zim	
https://www	Asenapine (s	Asenapine is	Asenapine is	Atypical anti	Saphris

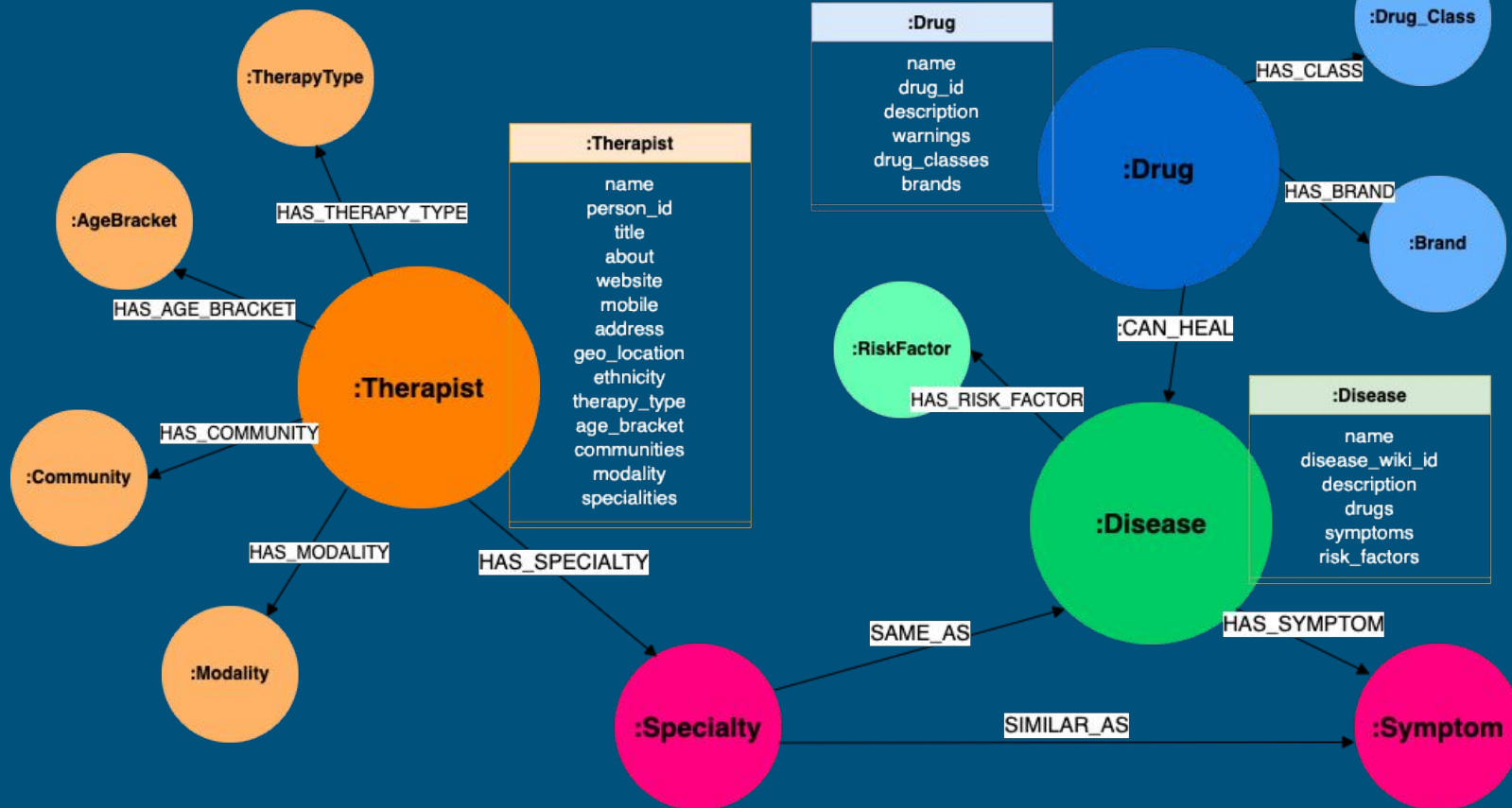
153 entries x 6 props



# Ontology



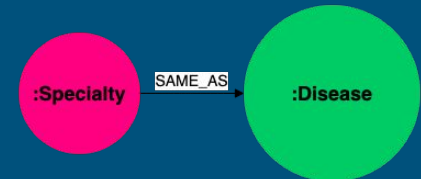
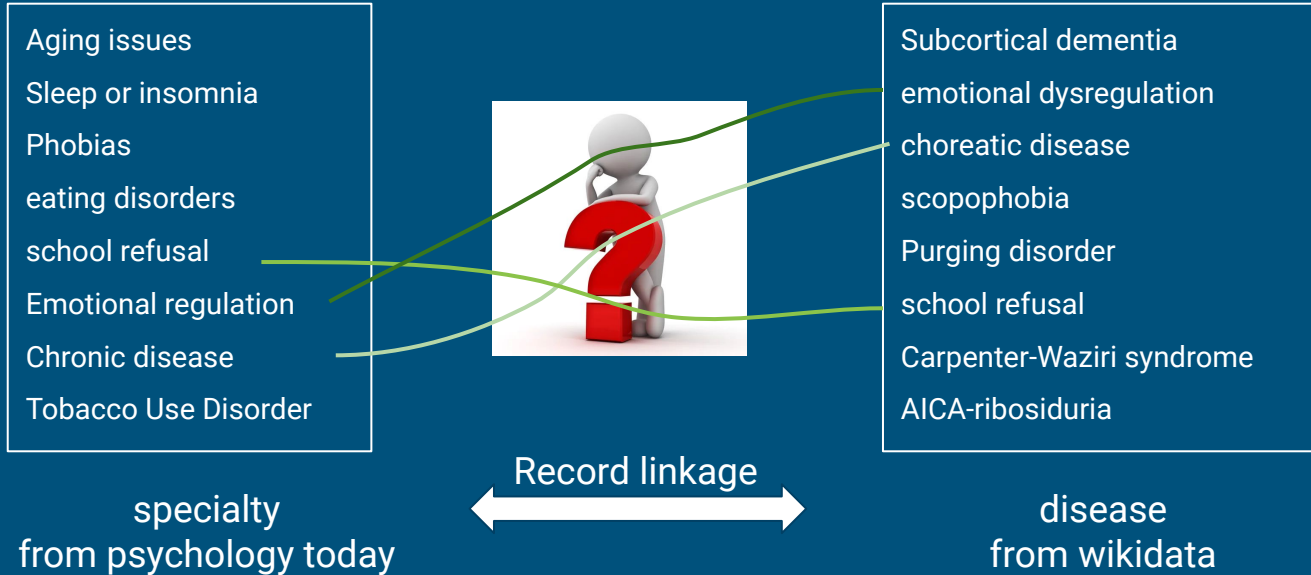
USC



# Technical Challenge : Record Linkage



USC



# Solution & Evaluation

Attempt: String matching

?



Levenshtein?  
Smith-waterman?  
Jaro?  
Jaccard?

Problem

- How to find the best threshold?
- Which string matching algorithm is the best one?



We need to build  
ground truth

**Jaro Example**

	D	I	X	O	N
D	1	0	0	0	0
I	0	1	0	0	0
C	0	0	0	0	0
K	0	0	0	0	0
S	0	0	0	0	0
O	0	0	0	1	0
N	0	0	0	0	1
X	0	0	0	0	0

$|x| = 5$   
 $|y| = 8$   
 $\text{max-distance} = \lceil (8/2) \rceil - 1 = 3$   
 $m = 4$   
 $t = 0$

**Jaccard Measure**

$B_x = \text{tokens}(x)$   
 $B_y = \text{tokens}(y)$

$$\text{jaccard}(x, y) = \frac{|B_x \cap B_y|}{|B_x \cup B_y|}$$

$\text{jaccard}(\text{dave}, \text{dav})$   
 $B_x = \{\#d, da, av, ve, e\# \}$   
 $B_y = \{\#d, da, av, v\# \}$   
 $\text{jaccard}(x, y) = 3/6$

alignment score matrix

$s(c_i, c_j)$

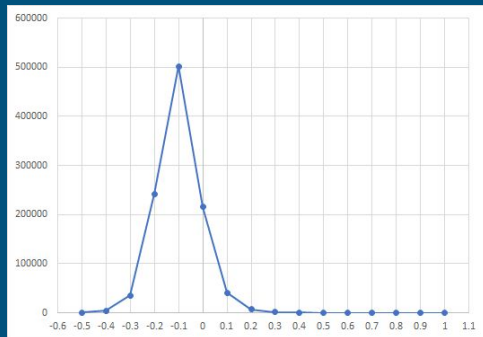
gaps

gap score

# Evaluation

## How to validate? Generate ground truth

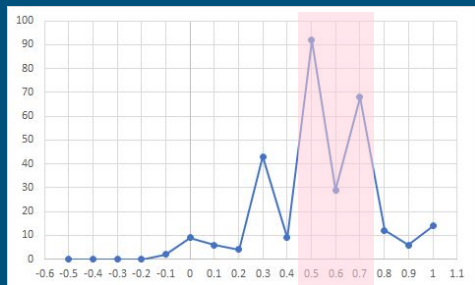
Step 1: All pairs (Needleman Wunsch)  
Low confidence pairs are useless for evaluation



Step 3: Set ground truth label

dating anxiety	death anxiety	0
down syndrome	Anton syndrome	0
panic disorder	tic disorder	0
obsessive-compulsive (OCD)	obsessive-compulsive disorder	1
agoraphobia	agoraphilia	1
eating disorder	pain disorder	0
eating disorder	writing disorder	0
imposter syndrome	Primrose syndrome	0
antisocial personality	dysocial personality	1
complex post traumatic stress disorder - CPTSD	complex post-traumatic stress disorder	1
adult autism spectrum disorder	autism spectrum disorder	1
suicidal ideation	Homicidal ideation	0
executive functioning	executive dysfunction	1
erectile dysfunction	executive dysfunction	0
adjustment disorder	adjustment disorder	1
bipolar disorder	bipolar II disorder	1

Step 2: Selected data  
(choose more percentage in 0.5~0.7)



Step 4: Experiment

String Matching Approach	precision	recall	F1 score	tp	fp	tn	fn
Levenshtein	0.8275862069	0.5853658537	0.6857142857	24	5	248	17
Needleman Wunsch	0.78125	0.6097560976	0.6849315068	25	7	246	16
Smith waterman	0.5185185185	0.6829268293	0.5894736842	28	26	227	13
Jaro winkler	0.8181818182	0.6585365854	<b>0.7297297297</b>	27	6	247	14
Jaccard	0.6896551724	0.487804878	0.5714285714	20	9	244	21
Hybrid jaccard	0.8181818182	0.4390243902	0.5714285714	18	4	249	23

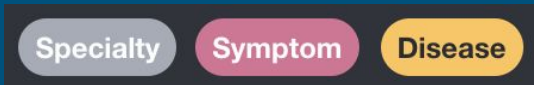
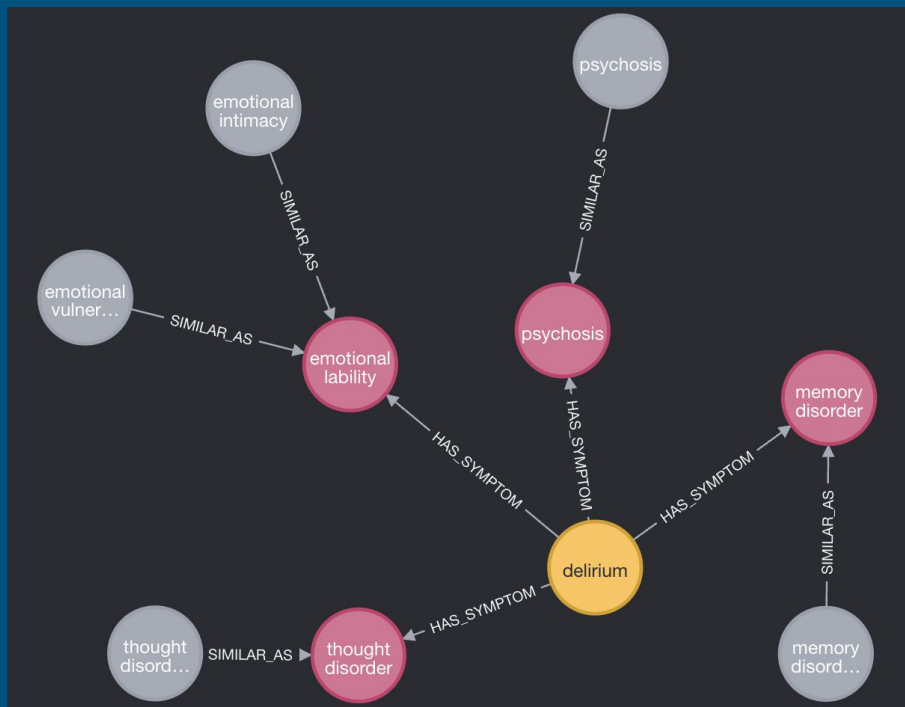
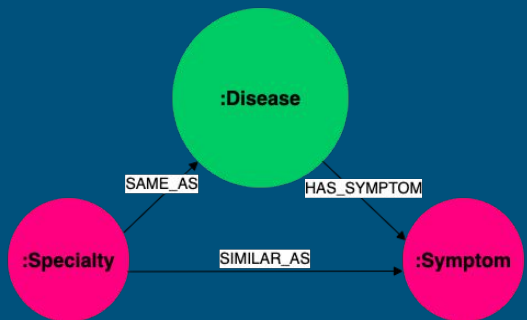
# Alternative

## Another Problem

- Matched pairs are sparse

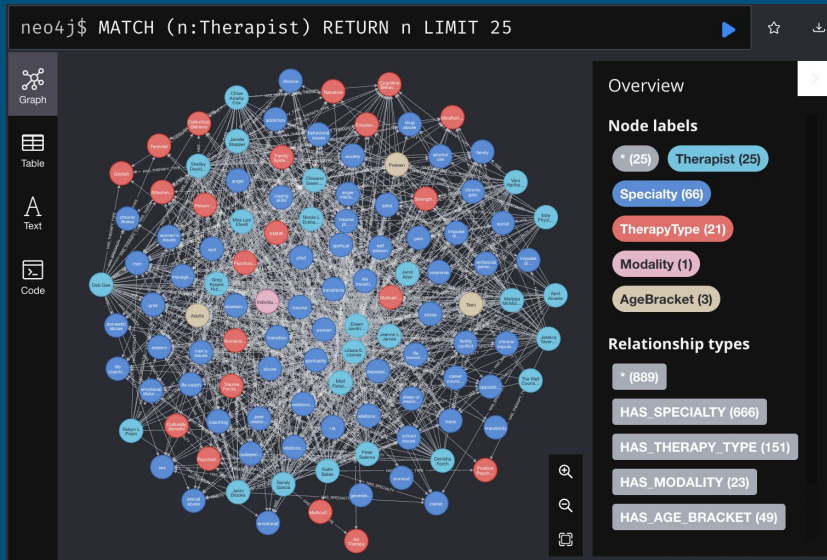
## Solution:

- Create linkage/relationships between symptoms(from wikidata) and specialty(from psychology today)

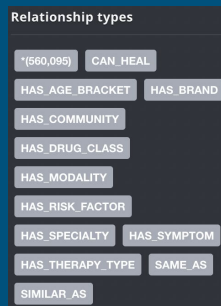
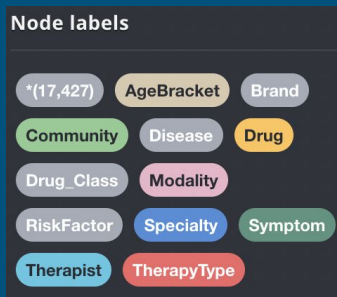




# Neo4j: Overview



- 17,427 nodes created
- 139,662 properties setted
- 560,055 relationships created
- around 1 hour to generate



# Neo4j: Application Scenarios

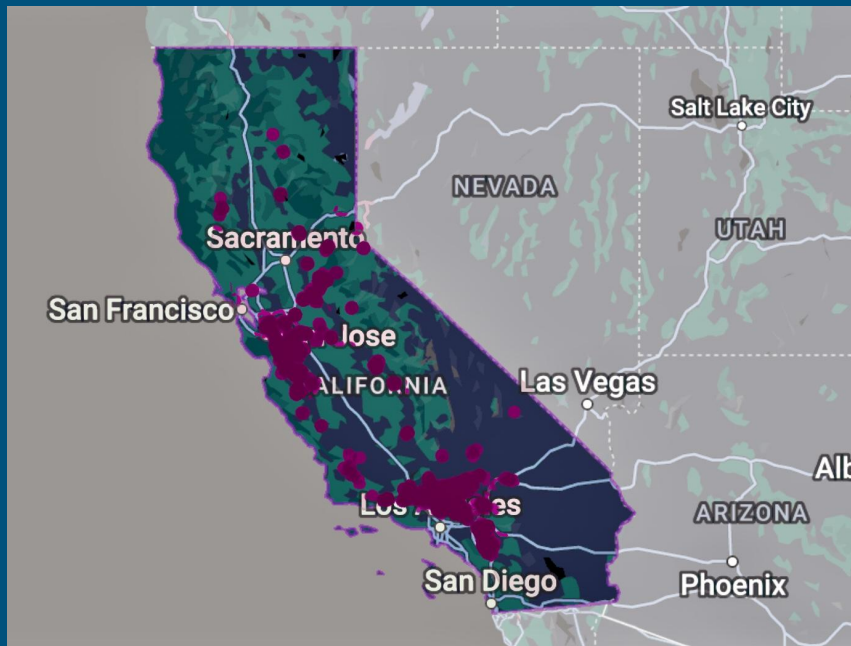
Q: I want to know therapists who can deal with insomnia around me.

**MATCH**

```
(t:Therapist)-[r1:HAS_SPECIALTY]->()-[r2:
SAME_AS]->(d:Disease{name:'insomnia'})
```

**RETURN** t.geo\_location

```
|-----|
| "t.geo_location" |
|-----|
| "32.8401623,-117.2740777" |
|-----|
```



1943 therapists found

# Neo4j: Application Scenarios

Q: I am rejection sensitive, I want to know which medicine might help this problem.

```
MATCH (s:Symptom{name:'rejection sensitive dysphoria'})<-[r1:HAS_SYMPTOM]-()-[r2:CAN_HEAL]->(d:Drug)RETURN d.name
```

Drug names:

Dexmethyphenidate, pemoline,  
Dextroamphetamine, Nortriptyline, Modafinil,  
Imipramine, Bupropion, Methylphenidate,  
Venlafaxine, Desipramine, Atomoxetine,  
Guanfacine, Clonidine



# Project Web-page – Basic Recommendation



Basic Recommendation

User Customized Recommendation

Disease: Major Depressive Disorder

Search

**Description**  
all-encompassing low mood accompanied by low self-esteem, and by loss of interest or pleasure in normally enjoyable activities

**Symptoms**  
fatigue, dysphoria, executive dysfunction, mental depression

**Risk Factor**  
family history, substance abuse, chronic condition

Please Input Address

Therapy Type

Ethnicity

Modality

Community

Age

Find Therapist

Name	Title	Address	Mobile	Specialities	Age	Modality	Community	Therapy Type	About	Website
Rachael E Barreto	Clinical Social Work/Therapist, LCSW	881 Fremont Ave Suite A3, Los Altos, 94024	(650) 397-7465	stress, anxiety, depression, relationship issues, codependency, coping skills, adhd, neurodiversit...	Adults	Individuals, Couples		Coaching, Cognitive Behavioral (CBT), Emotionally Focused, Interpersonal, Mindfulness-Based...	I am a licensed clinical social worker specializing in the treatment of adult...	
Cynthia Medina	Psychologist, PhD	415 N. San Mateo Drive, San Mateo, 94401	(650) 332-4688	child, anxiety, depression, education and learning disabilities	Teen, Children (6 to 10)				I am a licensed psychologist specializing in evaluations for children and ad...	<a href="#">website</a>

Search disease and related drugs

Filter conditions

Recommended therapists

# Project Web-page – Customized Recommendation



Basic Recommendation

User Customized Recommendation

Please tell us what troubles you have these days

I'm getting bored these days. Starting this Summer, it was getting harder for me to feel anything interesting. Sometimes I would be in a low temper and feel really emotional and upset for no reason. I can't stand this situation anymore, and I really need help.

Submit

Name	Title	Address	Mobile	Specialities	Age	Modality	Community	Therapy Type	About	Website
Nicole L Eckhardt	Counselor, LPCC, NCC	516 Villa Ave, Clovis, 93612	(559) 540-8186	child, stress, divorce, anxiety, peer relationships, depression, relationship issues, self esteem, school issues...	Adults, Teen, Preteen	Individuals, Couples, Family		Cognitive Behavioral (CBT), Eclectic, Family / Marital, Integrative, Interpersonal, Mindfulness...	I enjoy working with people in all stages of life, especially older teens, young adults...	<a href="#">website</a>
Mira Levi Elwell	Clinical Social Work/Therapist, LCSW	3911 Harrison Street, Oakland, 94611	(510) 298-3311	stress, anxiety, depression, self esteem, school issues, teen violence, emotional disturbance	Adults, Teen	Individuals, Family	Bisexual Allied, Gay Allied, HIV / AIDS Allied, Intersex Allied, Lesbian...		I am a LCSW with nearly a decade of social work experience in clinic, school...	<a href="#">website</a>

Users can describe their problems here

Recommended therapists

? How to recommend satisfied therapist?





# Technical Challenge – Recommend Satisfied therapist

Key problem: Diagnosis issue by user's description

? – Where do we get issue dataset?



Extract issues from specialties  
(from psychology today)

Problem of specialties data:

- Noisy data
- Too many types
- self-defined issue

Choose top 31 high frequency issues  
(Cover 71% issues)



Figure out a subset problem

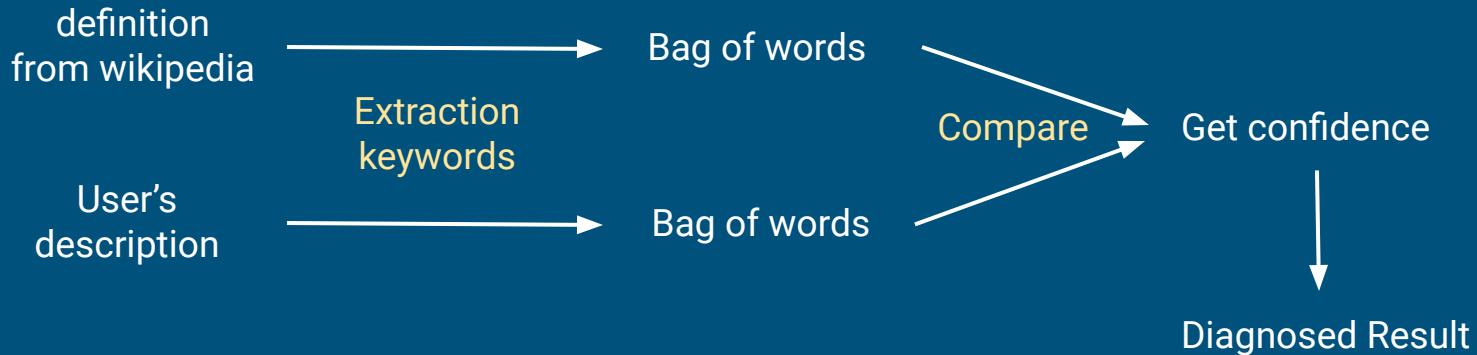
issue	count
anxiety	9295
depression	8892
relationship issues	8355
self esteem	7735
trauma and PTSD	7210
grief	6882
coping skills	6831
stress	6611
family conflict	6126
life transitions	6083
peer relationships	5156
parenting	5148
women's issues	5063
divorce	4543
codependency	4177
impulse control disorders	4057
marital and premarital	3877
behavioral issues	3707
child	3654
spirituality	3653
anger management	3500
addiction	3211
school issues	3194
emotional disturbance	3183
substance use	3040
sexual abuse	3027
life coaching	2999
adhd	2974
bipolar disorder	2565
infidelity	2483
suicidal ideation	2377

# Technical Challenge – Recommend Satisfied therapist



Key problem: Diagnosis issue by user's description

? – How to diagnose Psychological issue?



# Evaluation

? – How to build ground truth?



Collect description from friends.

Label data by psychology student.



Multiple label.

I'm getting bored these days. Starting this Summer, it was getting harder for me to feel anything interesting. Sometimes I would be in a low temper, and feel really emotional and upset for no reason. I can't stand this situation anymore, and I really need help.

I feel like I'm at a loss for my future every day. I want to do everything perfectly, but ultimately find myself just spinning in place with no progress.

Recently I could hardly fall asleep. Every time after I lay down, all negative things filled me. I feel stressed but don't know how to deal with it.

Behaviour issue  
Depression

Depression  
Grief

Stress  
Insomnia

# Evaluation

We use two metrics to evaluate our results.

$$\text{Recall}@R = \frac{\sum_{r=1}^R \mathbb{1}[(m(r) \in I_u)]}{\min(R, |I_u|)}$$

$$\text{DCG}@R = \sum_{r=1}^R \frac{2^{\mathbb{1}[(m(r) \in I_u)]} - 1}{\log(r + 1)}$$

Evaluation by Recall@10, Recall@5, Recall@3  
and DCG@10, DCG@5, DCG@3.

issue	confidence
depression	0.984615385
grief	0.861111111
behavioral issues	0.738636364
relationship issues	0.727272727
family conflict	0.673076923
divorce	0.615384615
infidelity	0.595238095
life transitions	0.434928329
women's issue	0.205882353
suicidal ideation	0.083333333

# Knowledge Graph Quality



## Data Completeness Analysis

Total Completeness

98.15%

therapists.csv	87.5%	mod_link.csv	100%	risk_link.csv	100%
therapy_type_df.csv	100%	specialities_df.csv	100%	drugs.csv	96.6%
therapy_type_link.csv	100%	spec_link.csv	100%	drug_disease_link.csv	100%
age_df.csv	100%	disease.csv	48%	class_df.csv	100%
age_link.csv	100%	symp_df.csv	100%	class_link.csv	100%
communities_df.csv	100%	symp_link.csv	100%	brand_df.csv	100%
com_link.csv	100%	symp_speclink.csv	100%	brand_link.csv	100%
modality_df.csv	100%	risk_df.csv	100%	person_spec_link.csv	100%

But What's Next?



# Knowledge Graph Quality

## - Fastness Check with Google

### Senario

My little sister's new friend gets an abnormal fatigue, usually she does nothing but feels tired and sleepy, I want to know why did she has these symptoms, and I want to find her a therapist who could help children like her.



### Our KG Solution

#### MATCH

```
(s.Symptom{name:"fatigue"})<-[r1:HAS_SY
MPTOM]-()-<[r2:HAS_RISK_FACTOR]->(r:
RiskFactor)RETURN r.name
```

#### MATCH

```
(s.Symptom{name:"fatigue"})<-[r1:SIMILAR
_AS]-()-<[r2:HAS_SPECIALTY]->(t:Therapis
t)-[r3:HAS_AGE_BRACKET]->(a:AgeBracke
t{name:"Children (6 to 10)"} ) RETURN
t.website
```

Get 4 factors and 3 recommended Therapists in 17ms

fast

VS

### Google Searching

slow

First search to find what disease may applied by given symptoms

Second search to find what might cause the disease

Third search on finding children-targeted therapists

But What if the therapist could not solve issues related to fatigue .....

? min

## - Task Completion Quality Check

### User Developed Task Check Feedback

Task	Result	Solvable
How many mental health problems are caused by smoking and social isolation?	6	Yes
Could you find any therapist who can help with bisexual teenagers' mental health problem?	Yes	Yes



Solving Rate

90%

Totally 10 Questions

# Conclusion

## Takeaway:

Knowledge graph is a powerful tool to well integrate multiple data sources, and especially knowledge graphs with great ontologies could be very convenient to trace detailed information.

Sequence-based string matching measures performed generally better than set-based measures for our dataset.

## Future work:

- Collect more data from different sources to expand database

- Test our knowledge graph relationships (especially the relationships between Symptom and Specialty) with more advanced instructs, and improve its quality by adjusting the string matching solution

- Collect more user description data to further polish the customized recommendation system