3.X

-

Reference Docs

ⓘ
- User Guide
- Human annotation with Protege
- Machine Annotations and Data Dictionary

Venn Diagram
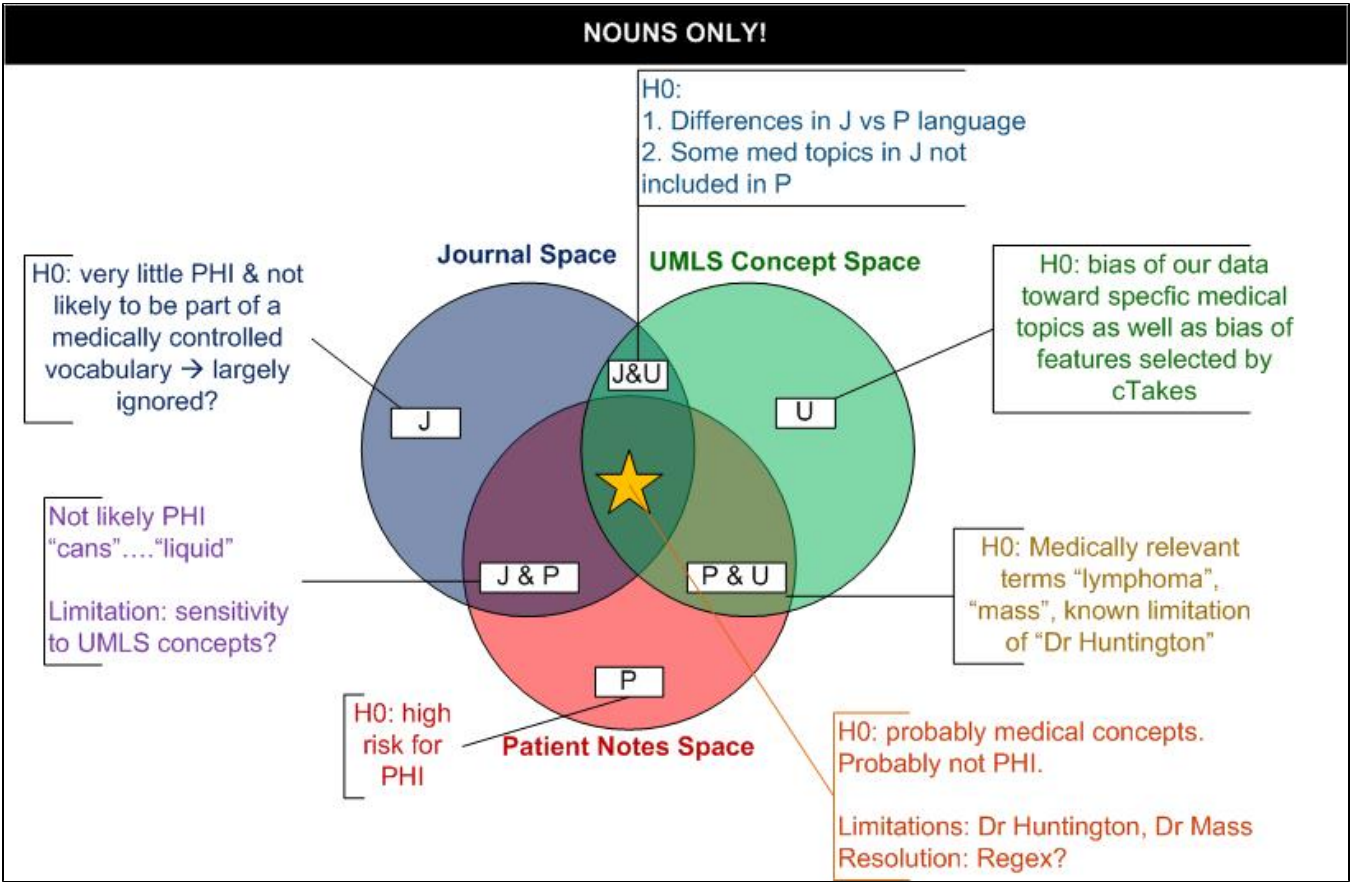
Feature Set Construction (Text words-> Lexical Features)

token	dict	regex	pub	pub freq	word	umls	umls freq	ni	class	prediction
the	0	0	1	115318		0	0		F	???
patient	0	0	1	620		1	20		F	
was	0	0	1	14639		1	6		F	
seen	0	0	1	562		1	1		F	
with	0	0	1	19400		1	3		F	
normal	0	0	1	871		1	20		F	
visit	0	0	1	39		1	4		F	

status	0	0	1	513	1	6		F
.	0	0	1	102847	0	0		F
patient	0	0	1	620	1	20		F
was	0	0	1	14639	1	6		F
checked	0	0	1	53	1	1		F
for	0	0	1	18770	1	5		F
kidney	0	0	1	60	1	15		F
failure	0	0	1	154	1	5		F
.	0	0	1	102847	0	0		F
patient	0	0	1	620	1	20		F
id	0	0	1	28	1	8		F
is	0	0	1	17343	1	7		F
12345	0	1	0	0	0	0		T
and	0	0	1	55187	1	4		F
name	0	0	1	51	1	10		F
is	0	0	1	17343	1	7		F
britt	1	1	0	0	0	0		T
fitch	1	1	1	1	0	0		T
.	0	0	1	102847	0	0		F

(i) 3.X is a new vision for the scrubber. As we approached diminishing returns for improving REGEX and whitelists/black lists, we have shifted towards a machine learning methods approach and learning from large bodies of medical information from publications and UMLS dictionaries.