

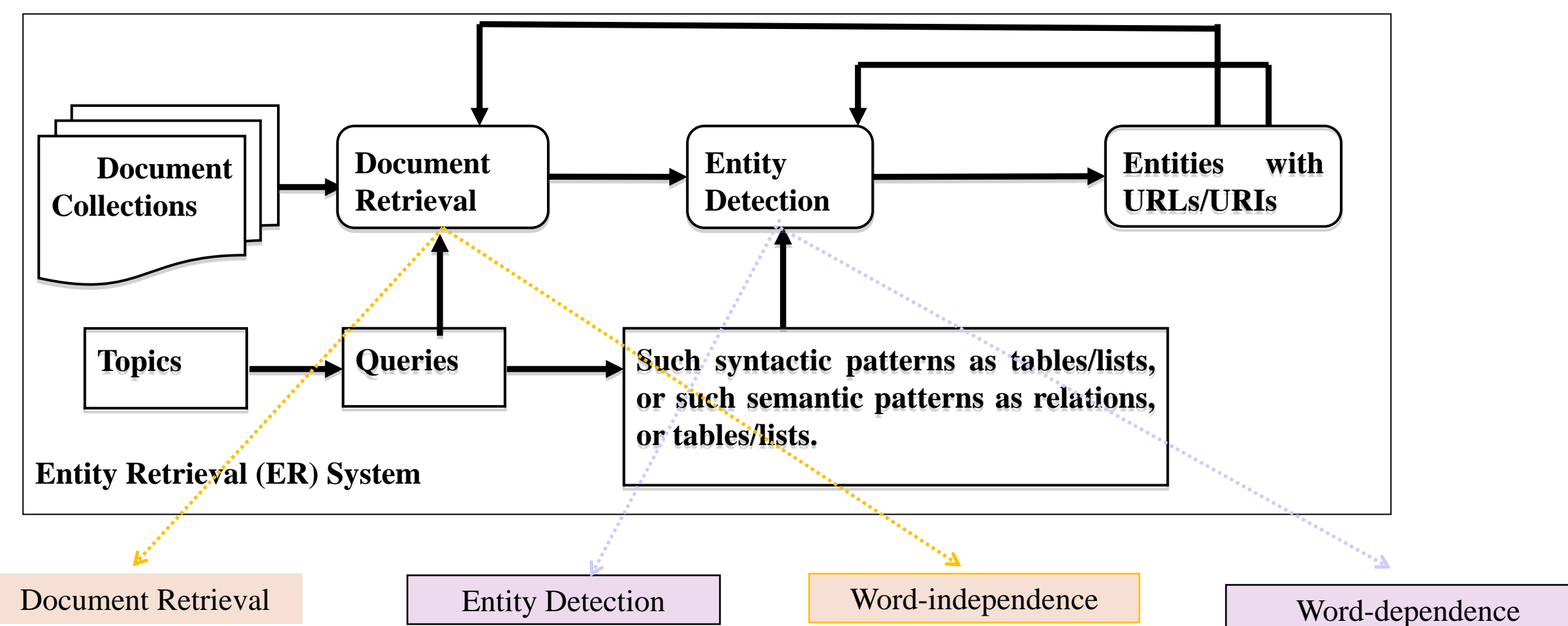


# Searching for Entities When Retrieval Meets Extraction

Qi Li, Daqing He University of Pittsburgh

## Motivation

- Review other teams' work
- Entity retrieval task is the combination task of document retrieval + entity detection + entity's URL/URI/s detections
- Retrieval on huge corpus for the finer-unit information -- entity
- System's performance and efficiency



### Advantages:

- Decompose the entity retrieval into two subtasks
- Adopt the state-of-the-art techniques into this work
- Improve the overall performances by studying each individual layers
- Study the relations between document retrieval and entity detections
- Entity's URL/URI mapping

$$\begin{aligned}
 & p(e | q, t_i) \\
 &= \sum_d \sum_c p(e, d, c | q, t_i) \\
 &= \sum_d \sum_c p(d | q, t_i) p(c | d, q, t_i) p(e | c, d, q, t_i) \\
 &= \sum_d p(d | q, t_i) \sum_c p(c | d, q, t_i) p(e | c, d, q, t_i) \\
 &\approx \sum_{d\_val} p(d\_val | q, t_i) \sum_{c\_val} p(c\_val | d\_val, q, t_i) p(e | c\_val, d\_val, q, t_i)
 \end{aligned}$$

Document Retrieval

Entity Detection

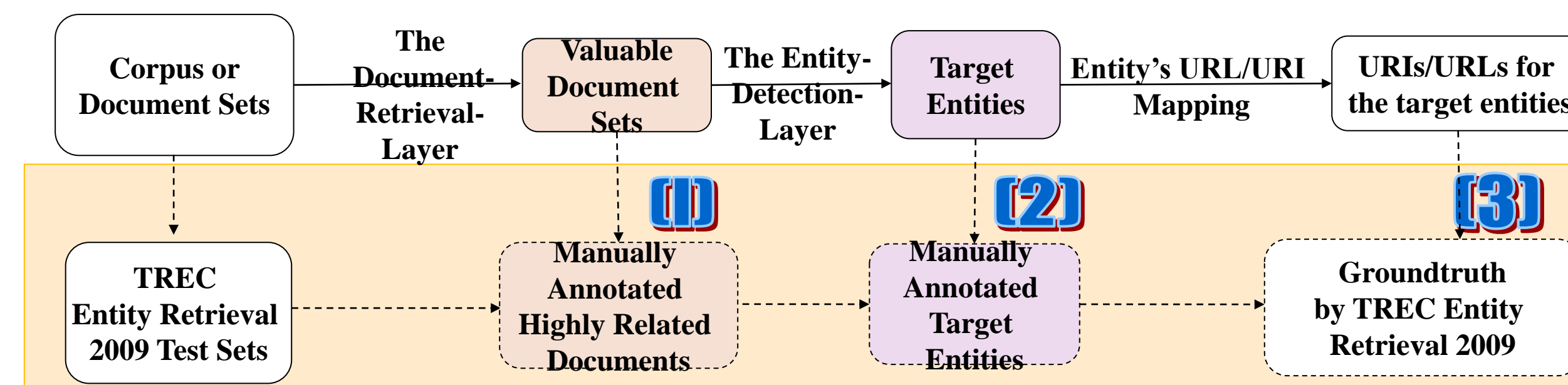
### Hint resources for 2010 topics

- Tables on the webpages: 28, 29, 40
- Lists on the webpages: 21, 24, 27(pdf), 31(cross several pages), 32, 33, 34 (pdf), 37, 38, 39, 41, 43(cross several pages), 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 59, 60, 61, 62, 66, 67, 68, 69, 70
- Tables on the Wiki pages: 22, 23, 29, 33, 36, 38, 42(category)
- Lists on the Wiki pages: 24, 25, 28, 51, 63, 64,
- Sentences on the webpages/wiki pages: 30, 42, 44, 45, 65
- Sentences scattering across several webpages: 26, 46, 58
- None: 35

## Math

## Evaluation

### Entity Retrieval Tasks



Topic :  
Products of Medimmune, Inc.

### Valued Documents:

[http://www.medimmune.com/about\\_us\\_products.aspx](http://www.medimmune.com/about_us_products.aspx)  
<http://en.wikipedia.org/wiki/MedImmune>

### Answer Entities:

- Synagis  
- FluMist  
- Ethyol

### GTH homepages:

- 5-HPclueweb09-en0000-27-129352  
<http://baltimore.bizjournals.com/baltimore/stories/2009/01/05/daily20.html> -- News  
- 5-HPclueweb09-en0006-42-198412  
<http://www.ethyol.com/> -- Products Homepage  
- 5-HPclueweb09-en0008-26-393002 <http://www.medimmune.com/>  
-- Company Homepage  
- 5-HPclueweb09-en0008-26-393062  
<http://www.medimmune.com/about/history.asp> -- Company Introduction page  
- 5-HPclueweb09-en0008-26-393282  
<http://www.medimmune.com/products/flumist/index.asp> -- Company Introduction page  
- 5-HPclueweb09-en0008-26-393302  
<http://www.medimmune.com/products/synagis/index.asp> -- Company Introduction page

Topic ID	Entity Type	# of Doc	# of Wiki Doc	# of Entities	# of Homepages	Topic ID	Entity Type	# of Doc	# of Wiki Doc	# of Entities	# of Homepages
1	Organization	1	1	4	10	11	Organization	2	0	8	8
2	Person	1	0	12	1	12	Organization	2	1	32	13
3	Person	0	0	0	1	13	Product	2	1	2	4
4	Organization	3	1	8	5	14	Person	1	1	36	4
5	Product	3	1	3	8	15	Organization	2	1	12	9
6	Organization	2	1	4	4	16	Organization	1	1	11	9
7	Organization	1	1	4	33	17	Person	1	1	61	18
8	Product	2	1	41	41	18	Person	2	1	14	3
9	Person	2	1	9	13	19	Organization	3	2	4	3
10	Organization	2	1	9	2	20	Organization	2	1	27	4