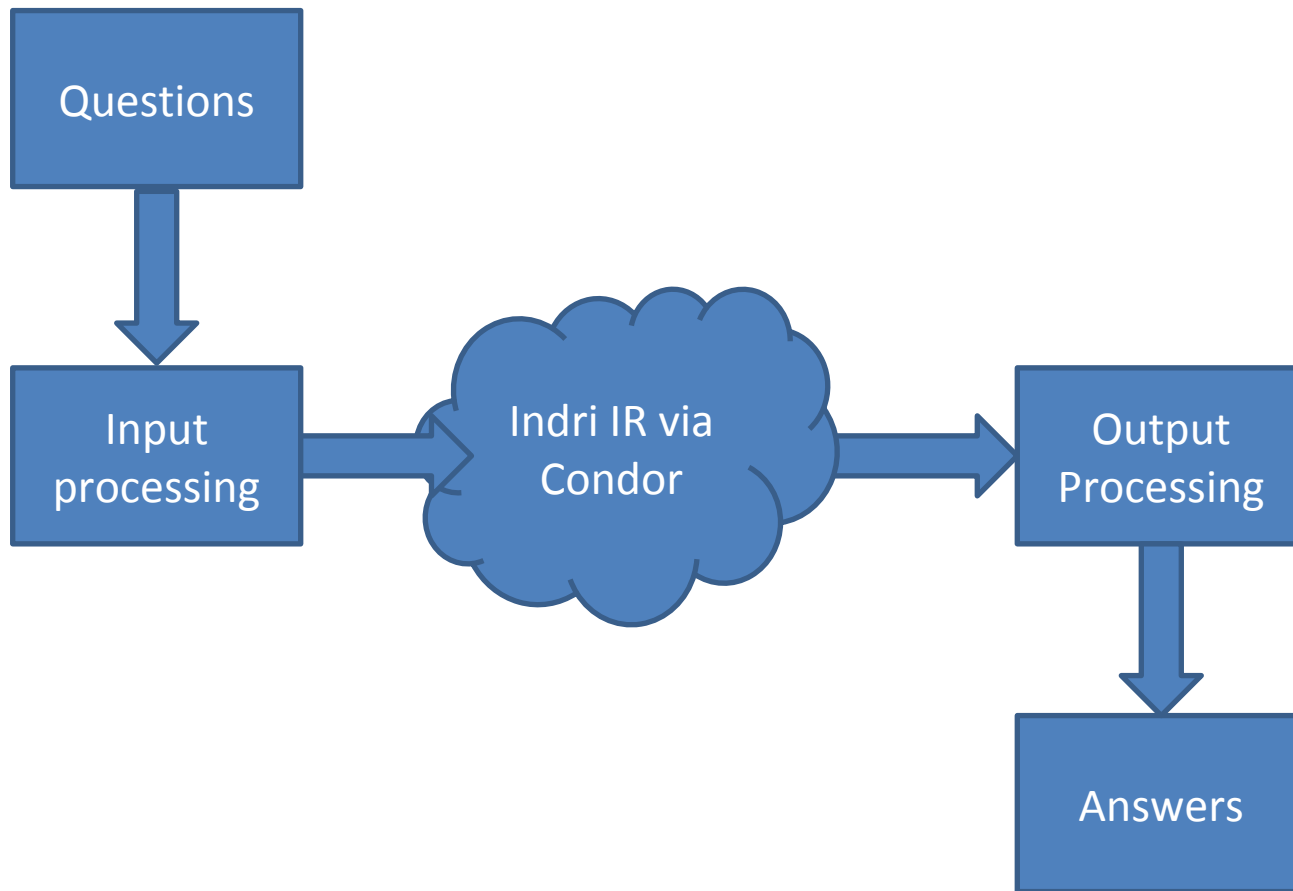# LING 573 D3
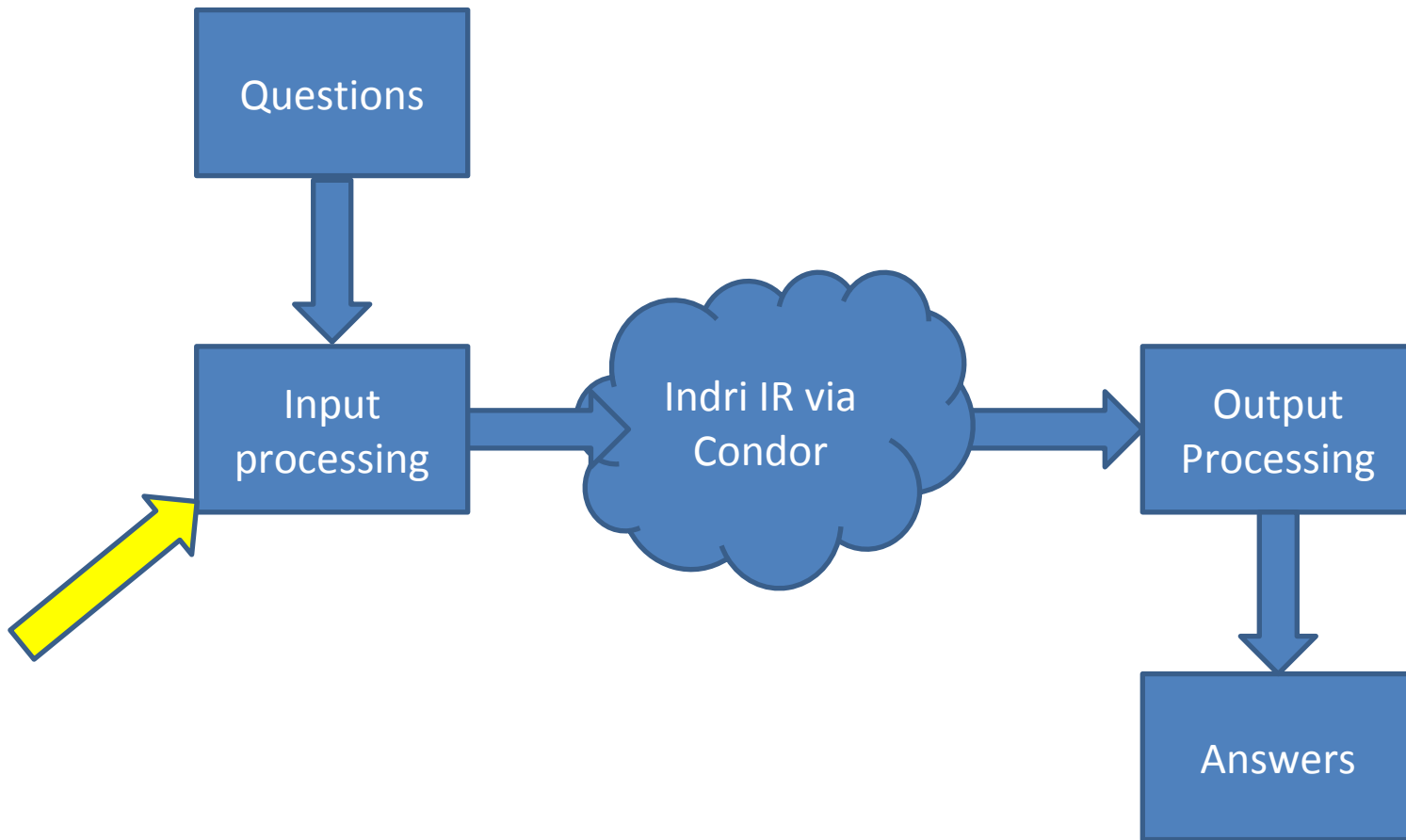# Query Expansion with Deep Processing

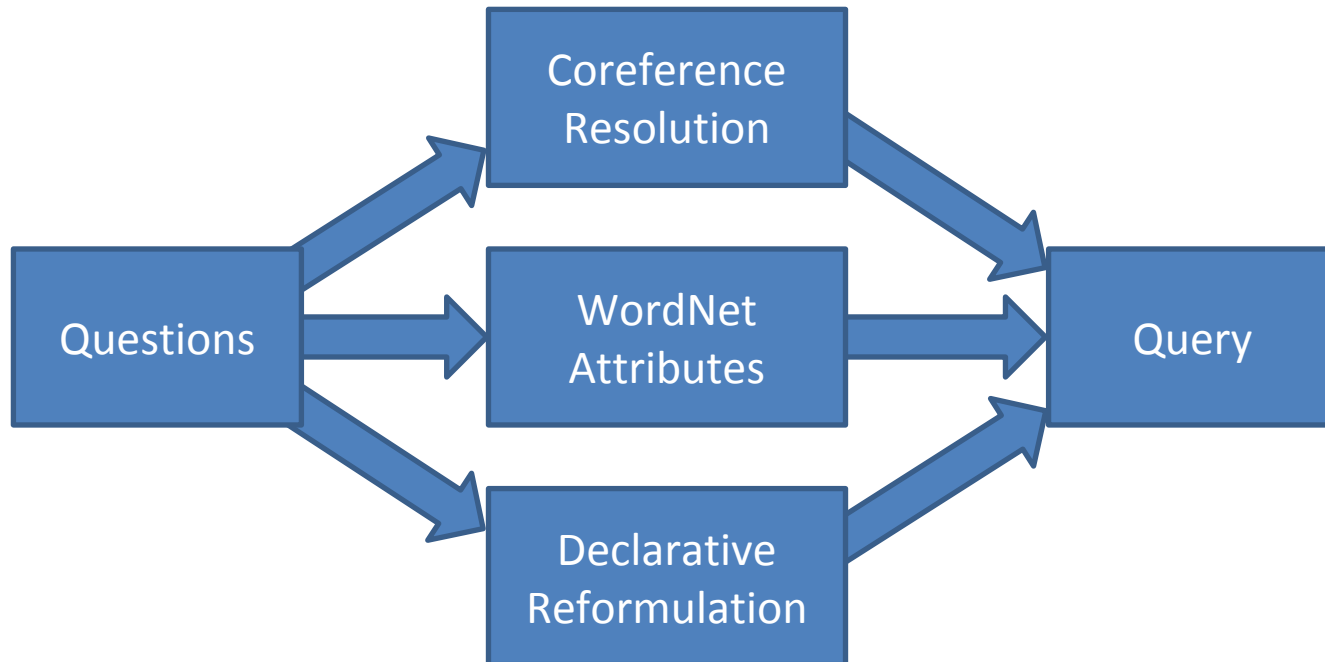Melanie Bolla, Woodley Packard, and T.J. Trimble
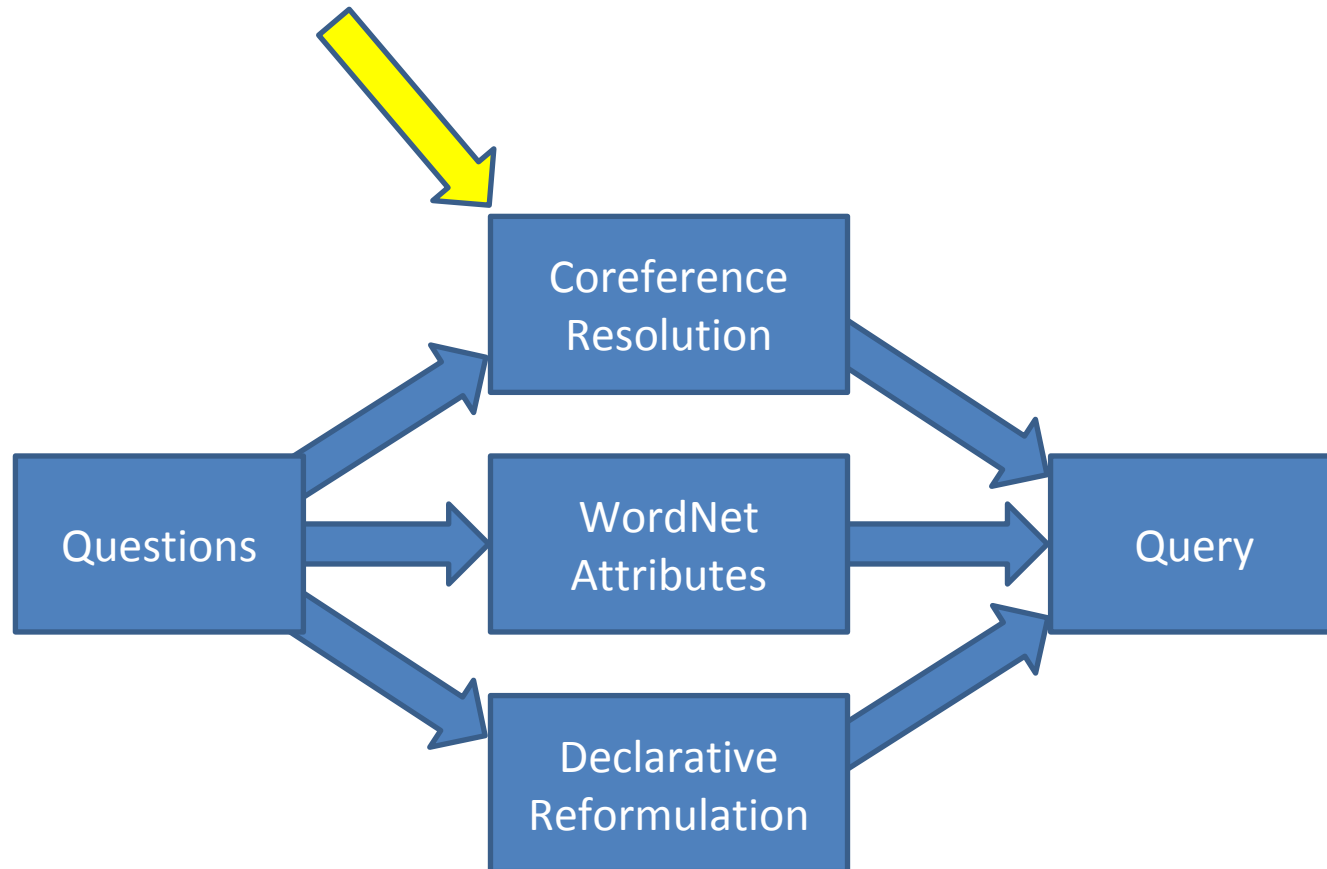
System Architecture

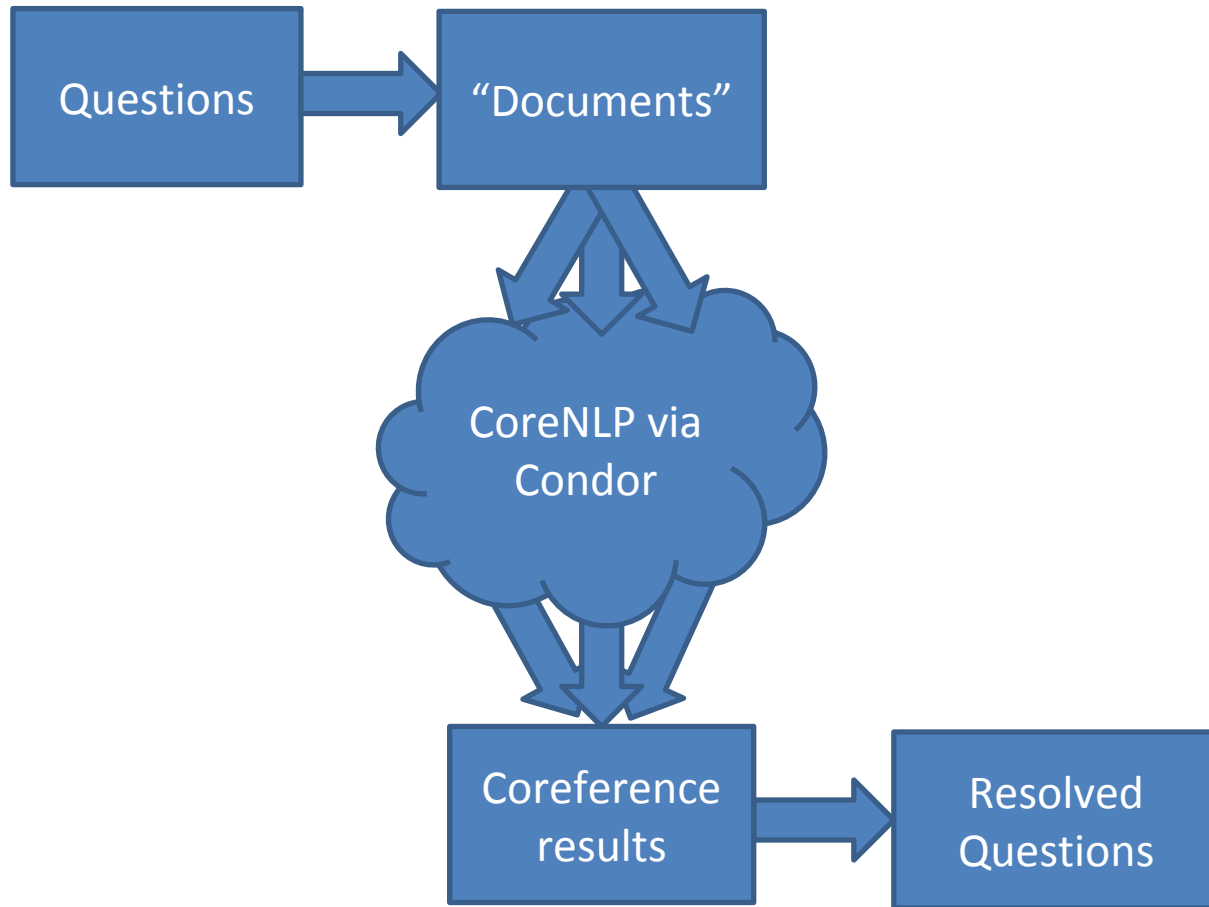# System Architecture

Input Processing

Input Processing

Coreference Resolution

- Intuition: Replace pronominal or underspecified references with antecedent
  - Do some clean up
- System: Stanford CoreNLP dcoref
  - Rule based sieve architecture for coreference resolution
- Implementation: Parallelization via Condor
- Improvements!

Coreference Resolution

Coreference Resolution

- Document
  - Target + question series
  - Coreference resolution is done over document

Bing Crosby. What was his profession? For which movie did he win an Academy Award? What was his nickname? What is the title of his all-time best-selling record? He is an alumnus of which university? How old was Crosby when he died?

→

Bing Crosby. What was Bing Crosby's profession? For which movie did he win an Academy Award? What was Bing Crosby's nickname? What is the title of Bing Crosby's all-time best-selling record? He is an alumnus of which university? How old was Crosby when he died?

Coreference Resolution

- Query Formulation:
  - Get replacements from dcoref
  - Do replacements over question file, with some additional cleaning (possessives, etc.)
  - Submit to Indri using #4(q)

Coreference Resolution

- Results:
  - Initial Results:
    - Baseline:
      - Lenient: 0.2390; Strict: 0.1525
    - Coref:
      - Lenient: 0.2013; Strict: 0.1339

Coreference Resolution

- Results:
  - Initial Results:
    - Baseline:
      - Lenient: 0.2390; Strict: 0.1525
    - Coref:
      - Lenient: 0.2013; Strict: 0.1339
- -_-`

Coreference Resolution

- Error Analysis:
  - Problematic resolutions:
    - `What is Crosby's nickname?`
    - `What is Crosby's wife's name?`
      - `-> What is What is Crosby's nickname's wife's name?`
  - Due to overzealous resolution in the face of impaired punctuation
  - Not very good regex replacement

Coreference Resolution

- Fixes (post-deadline):
  - Constrain replacements to only "the best"
    - extraneous determiner additions
    - make sure possessives line up right
    - enforce only adding content
    - etc.
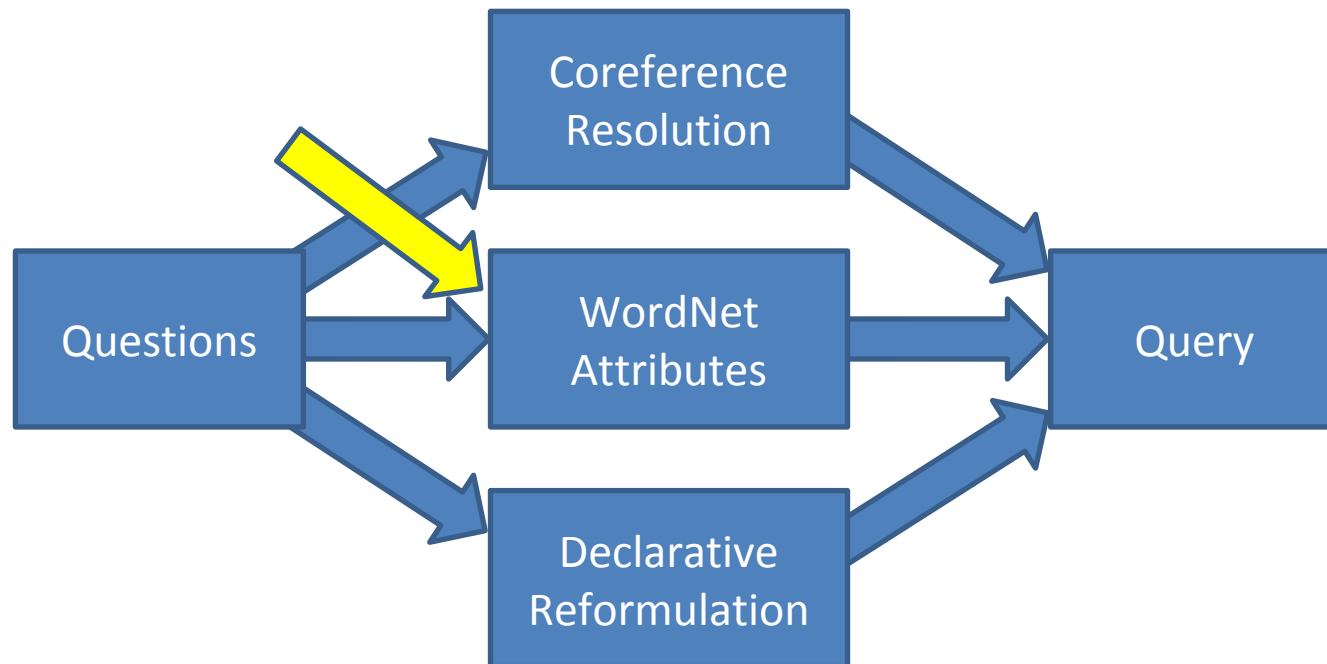 - On devtest: reduction in replacement candidates from about 160 to 72

Coreference Resolution

- Results:
  - Baseline: Lenient: 0.2390; Strict: 0.1525
  - Coref: Lenient: 0.2013; Strict: 0.1339
  - Baseline Improved:
    - Lenient: 0.2618; Strict: 0.1813
  - Coref Improved (post-deadline):
    - Lenient: 0.2780; Strict: 0.1868

Coreference Resolution

- Future Work:
  - What if coreference fed into declaratives?
    - Where did Moon play in college?
    - Where did Warren Moon play in college?
    - Warren Moon played in college.

Input Processing

WordNet Related Nouns

- Insert "related nouns" of adjectives in WordNet into bag of word query

- Intuition: "how tall" -> "height"

  - Initial drop in score

    - Baseline: Lenient: 0.2390; Strict: 0.1525

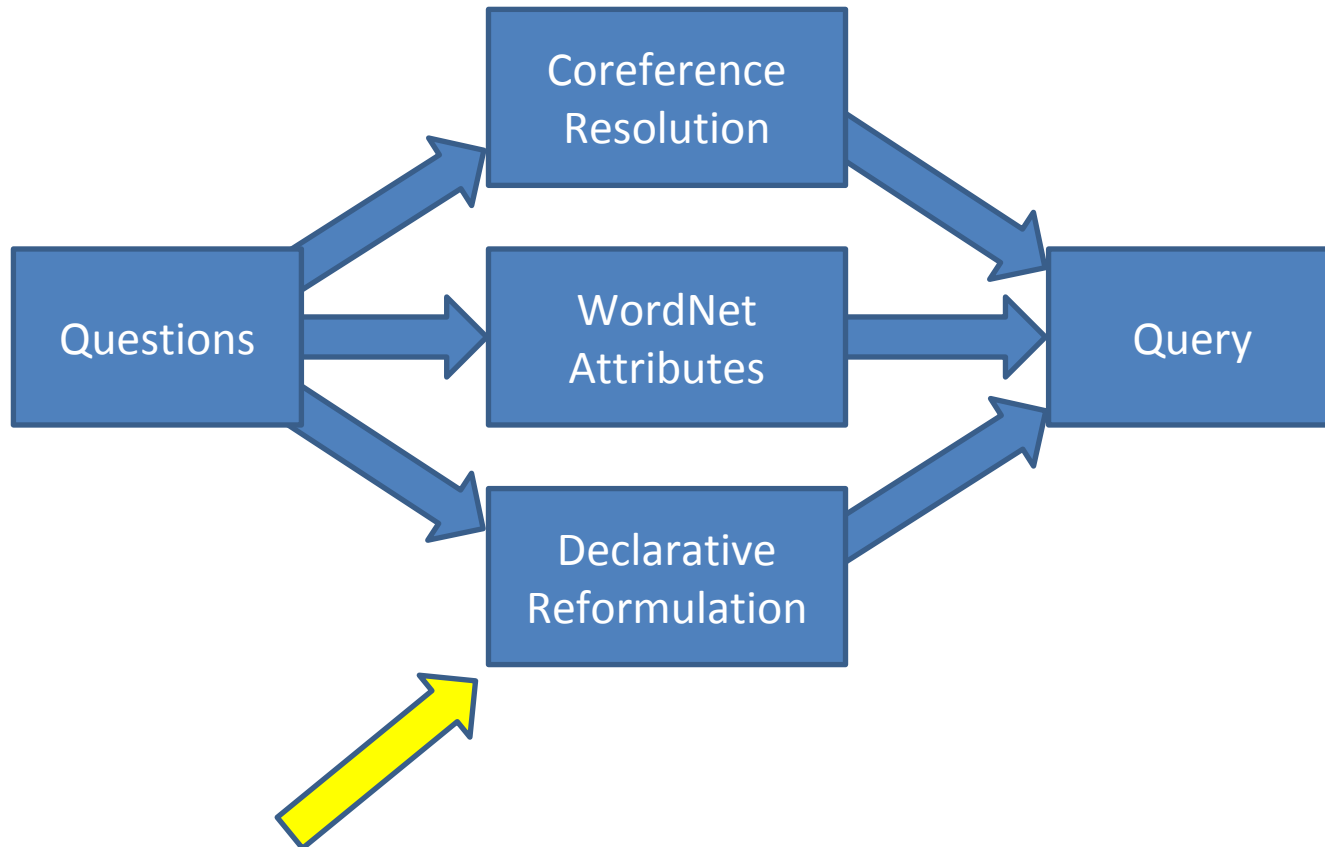   - Initial: Lenient: 0.2278; Strict: 0.1512

WordNet Related Nouns

- Error Analysis:
  - Some words had terrible attributes:
    - "current" -> "currentness, currency, up-to-dateness"
    - "other" -> "otherness, distinctness, separateness"
    - **"many" -> "numerousness, numerosity, multiplicity"**

WordNet Related Nouns

- Removed "many":
  - Baseline:
    - Lenient: 0.2390; Strict: 0.1525
  - Initial:
    - Lenient: 0.2278; Strict: 0.1512
  - Removed "many":
    - Lenient: 0.2378; Strict: 0.1563
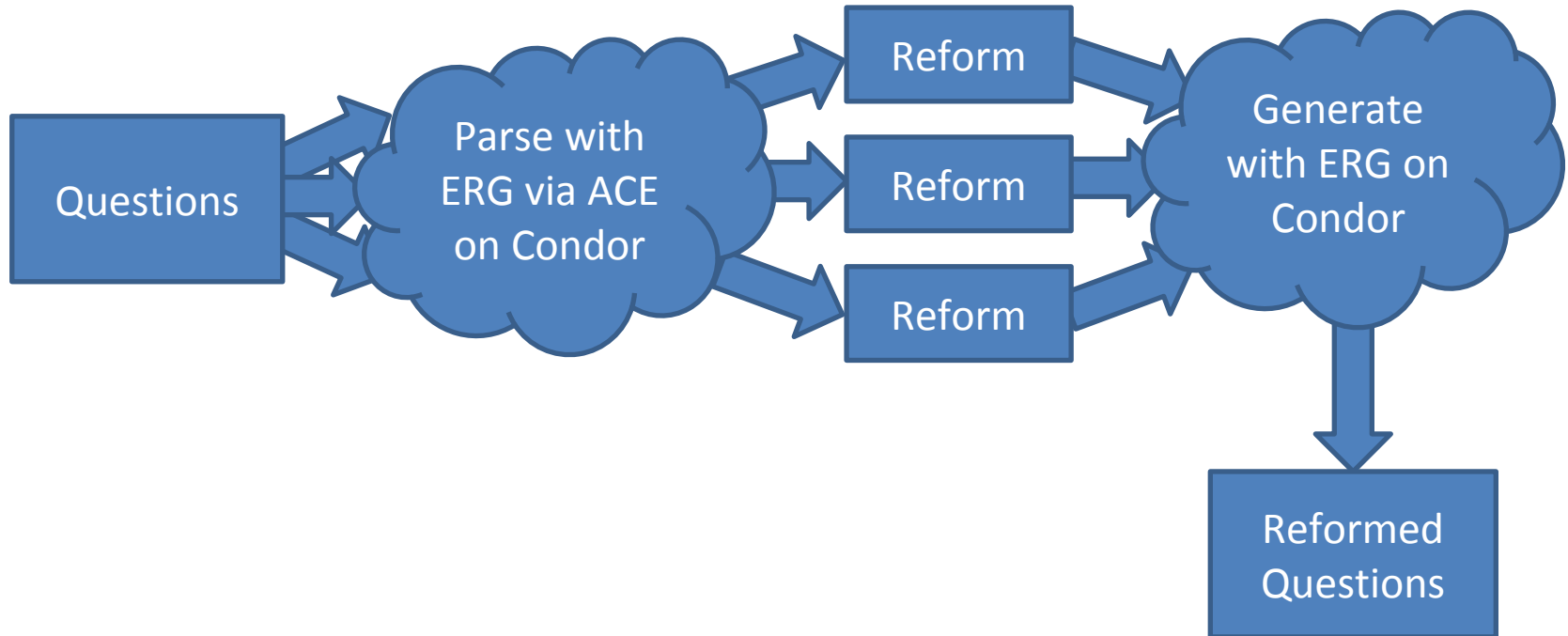
Input Processing

Declarative Reformulation

- Intuition: documents have statements, not questions; shallow reformulation stinks
- Declarative Reformulation using the ERG
  - Parse question into flat semantic representation, MRS
  - Fiddle with MRS
  - Generate with ERG
- Improvements!

Declarative Reformulation

- Input:
  - What position did Moon play in professional football?
  - Where did Moon play in college?
- Output:
  - A position did moon play in professional football.
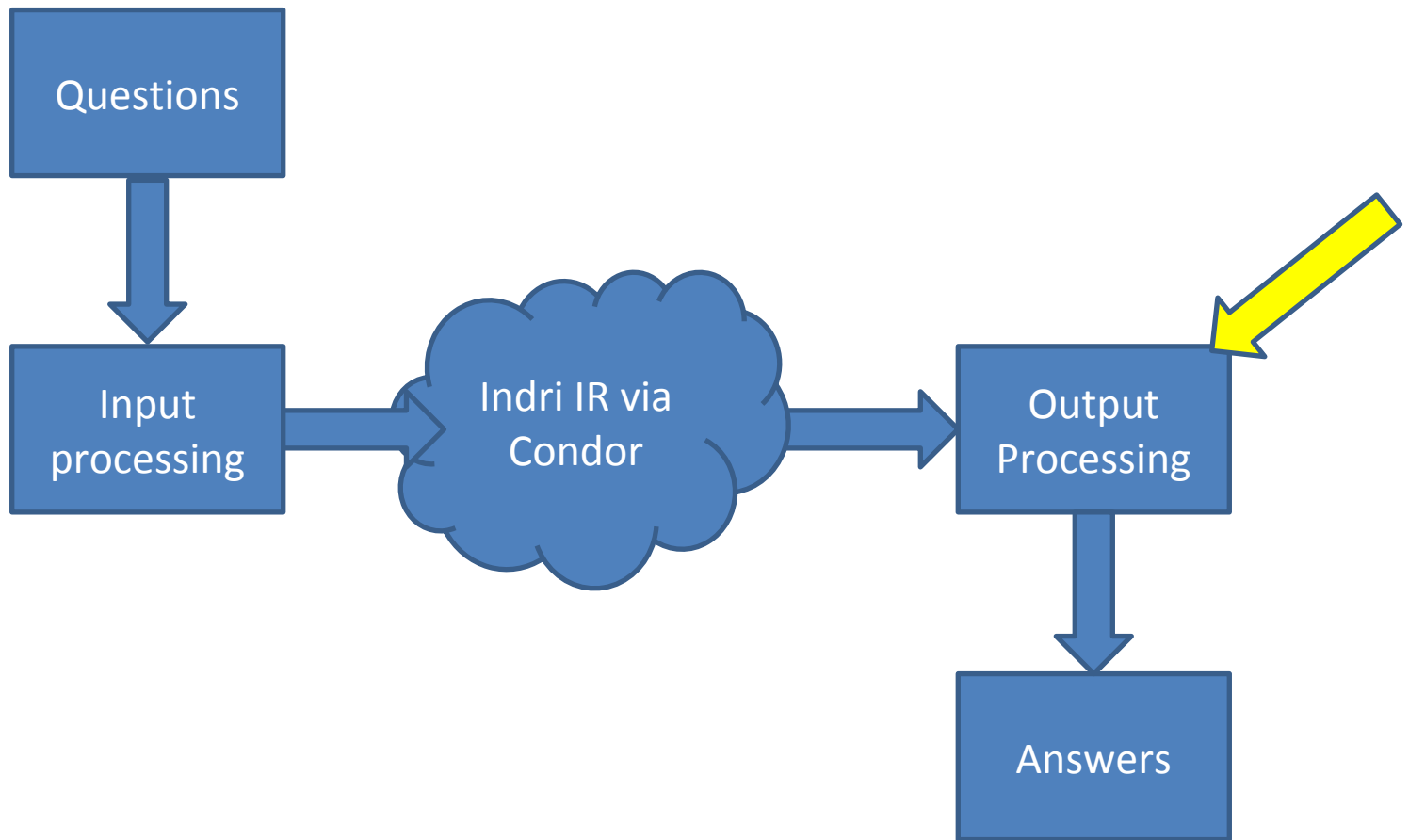 - Moon played in college.

Declarative Reformulation

Declarative Reformulation

- Baseline:
  - Lenient: 0.2618; Strict: 0.1813
- Declaratives:
  - Lenient: 0.2695; Strict: 0.1905

System Architecture

Questions

Input processing

Indri IR via Condor

Output Processing

Answers

Answer Processing

- Choosing better snippets
  - Starting from the center of the document seemed to work the best
    - This might be overfitting…
  - Baseline:
    - Lenient: 0.2390; Strict: 0.1525
  - Improvement:
    - Lenient: 0.2695; Strict: 0.1905

Answer Processing

- Remove HTML
  - 2 lines of code with NLTK
  - Baseline:
    - Lenient: 0.2621; Strict: 0.1835
  - Improvement:
    - Lenient: 0.2642; Strict: 0.1881

MRS matching

- Match question to answer based on MRS graph structure
  - Big improvement!
  - Baseline:
    - Lenient: 0.2695; Strict: 0.1905
  - MRS-matching:

    Lenient: 0.3263; Strict: 0.2452
  - Post-deadline:

    Lenient: 0.3317; Strict: 0.2564

Results (devtest)

Test	Lenient Score	Strict Score	IR Recall
Baseline	0.1319	0.0753	?
Baseline Improved (B)	0.2618	0.1813	67.5 / 55.6
B + Declarative (D)	0.2695	*0.1905*	68.4 / 57.1
B + WordNet Attributes (W)	0.2545	0.1743	66.5 / 54.6
B + Coreference (C)	*0.2780*	0.1868	?
D3: B + D + W	**0.2622**	**0.1835**	**67.5 / 56.1**
B + W + C	0.2706	0.1853	?
B + D + W + C	0.2642	0.1881	?

Results (devtest)
… with MRS matching

Test	Lenient Score	Strict Score
Baseline Improved (B)	0.3209	0.2379
B + Declarative (D)	0.3263	0.2452
B + WordNet Attributes (W)	0.3216	0.2398
Baseline + Coreference (C)	0.3343	0.2445
D3: B + D + W	**0.3269**	**0.2471**
Post-deadline: B + D + W + C	*0.3453*	*0.2565*

Issues

- Indri
  - Finding the best/proper Indri Query Language operators
- WordNet
  - WSD, weird relationships
- Coreference
  - Match happy system

Successes

- Taking 250 characters from the middle of the snippet
- Constraining Coreference Resolution
- Declarative Reformulation
- HTML cleaning
- MRS based matching

Influential Related Reading

- ERG and MRS: Copestake 2000, Copestake 2002, Flickinger 2003, Copestake 2005
- WordNet: ? class 10 slide 6
- Coreference Resolution: Raghunathan et al., 2010, etc.
- Class reading on Indri: http://sourceforge.net/p/lemur/wiki/Home/