Passage Scoring for Question Answering using Syntactic-Semantic Graphs

Soft Matching Extensions of LFACS

Annemarie Friedrich

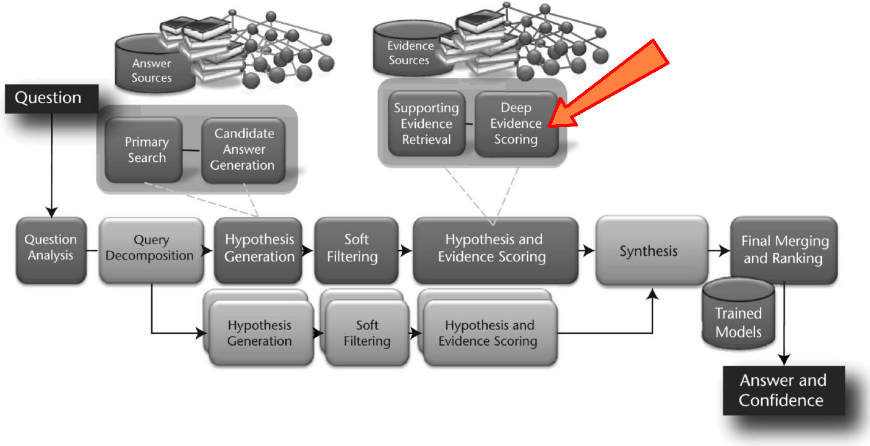
Department of Computational Linguistics, Saarland University
IBM T.J. Watson Research Center, NY

August 2nd, 2011

Outline

Outline

DeepQA Architecture



Supporting Passage Retrieval

FOCUS

QUESTION: *Thallium is said to look like **this element**.*

CANDIDATE ANSWER: ***lead***

Query Generation

$\Downarrow$ ***Supporting Passage Retrieval***

PASSAGE: *Thallium is a metallic element that looks like **lead**.*

CANDIDATE ANSWER

Supporting Passage Retrieval

FOCUS

QUESTION: *Thallium is said to look like **this element**.*

CANDIDATE ANSWER: ***lead***

Query Generation

$\Downarrow$ ***Supporting Passage Retrieval***

PASSAGE: *Thallium is a metallic element that looks like **lead**.*

CANDIDATE ANSWER

Deep Evidence Scoring

Task of Passage Scorers:

Determine degree to which passages justify candidate answer.

Question

The Neckar river begins its 228-mile course in ***this region*** *aka the Schwarzwald.*

Justification

Neckar, river, 228 miles long, rising in the ***Black Forest****.*

Partial Justification

Another main tributary of the Rhine is the Neckar, which drains the ***Black Forest*** *and the Swabian Alb. Running 228 miles in length, this meandering river is celebrated for its scenery and charm.*

Deep Evidence Scoring

Task of Passage Scorers:

Determine degree to which passages justify candidate answer.

Question

The Neckar river begins its 228-mile course in ***this region*** *aka the Schwarzwald.*

Justification

Neckar, river, 228 miles long, rising in the ***Black Forest****.*

Partial Justification

Another main tributary of the Rhine is the Neckar, which drains the ***Black Forest*** *and the Swabian Alb. Running 228 miles in length, this meandering river is celebrated for its scenery and charm.*

Deep Evidence Scoring

Task of Passage Scorers:

Determine degree to which passages justify candidate answer.

Question

*The Neckar river begins its 228-mile course in **this region** aka the Schwarzwald.*

Justification

*Neckar, river, 228 miles long, rising in the **Black Forest**.*

Partial Justification

*Another main tributary of the Rhine is the Neckar, which drains the **Black Forest** and the Swabian Alb. Running 228 miles in length, this meandering river is celebrated for its scenery and charm.*

Deep Evidence Scoring

Question

*The Neckar river begins its 228-mile course in **this region** aka the Schwarzwald.*

No Justification

*Sulz am Neckar is located right between the lovely Swabian Alb and the mystic **Black Forest**.*

Inter-annotator agreement

- 243 questions
- Cohen's $\kappa = 0.67$ (substantial)
- most disagreement: yes $\Leftrightarrow$ partial

Deep Evidence Scoring

Question

The Neckar river begins its 228-mile course in ***this region*** *aka the Schwarzwald.*

No Justification

Sulz am Neckar is located right between the lovely Swabian Alb and the mystic ***Black Forest***.

Inter-annotator agreement

- 243 questions
- Cohen's $\kappa = 0.67$ (substantial)
- most disagreement: yes $\Leftrightarrow$ partial

Passage Term Match (PTM)

Example

Thallium is said to **look like** this element.
Thallium is a metallic **element** that **looks like** lead.

- Counts how many question terms are matched in passage.
- PTM score = decaying sum of scores for all passages.

[Murdock et al., 2011]

Outline

Logical Form Answer Scorer (LFACS)

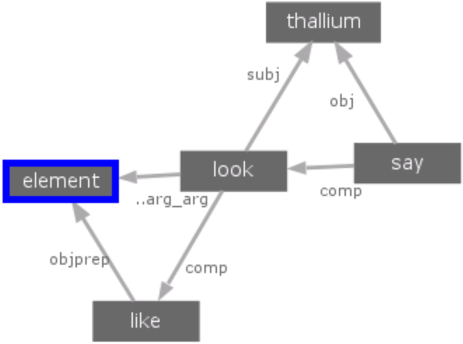
- Syntactic-Semantic Graphs (= Logical Forms)
- Term Matchers
- Graph Alignment with special attention to focus & candidate answer

[Murdock et al., 2011]

Syntactic-Semantic Graphs

- Slot Grammar (grammatical relations)
  `subj, obj, iobj, ndet,...`
- Shallow Semantic Relations
  `instanceOf, theme, experiencer`
- Derivational Morphology
  `nobj, obj` → `dm_obj_arg`
- Deep Semantic Relations
  `actorInOpus, bornWhen, nationalityOf,...`
- Coreference (Anaphora resolution)
- Predicate Argument Structure

[McCord et al., 2011]

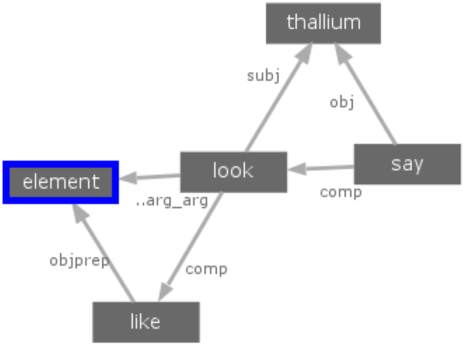


Thallium is said to look like this element.

Syntactic-Semantic Graphs

- Slot Grammar (grammatical relations)
  `subj, obj, iobj, ndet,...`
- Shallow Semantic Relations
  `instanceOf, theme, experiencer`
- Derivational Morphology
  `nobj, obj` → `dm_obj_arg`
- Deep Semantic Relations
  `actorInOpus, bornWhen, nationalityOf,...`
- Coreference (Anaphora resolution)
- Predicate Argument Structure

[McCord et al., 2011]

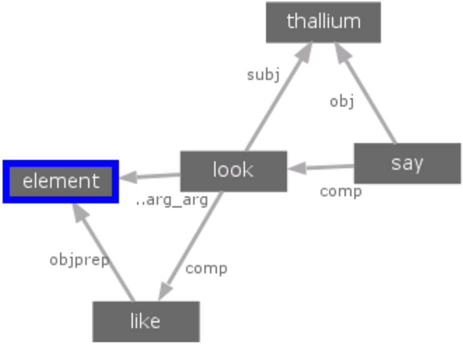


Thallium is said to look like this element.

Syntactic-Semantic Graphs

- Slot Grammar (grammatical relations)
  `subj, obj, iobj, ndet,...`
- Shallow Semantic Relations
  `instanceOf, theme, experiencer`
- Derivational Morphology
  `nobj, obj` → `dm_obj_arg`
- Deep Semantic Relations
  `actorInOpus, bornWhen, nationalityOf,...`
- Coreference (Anaphora resolution)
- Predicate Argument Structure

[McCord et al., 2011]

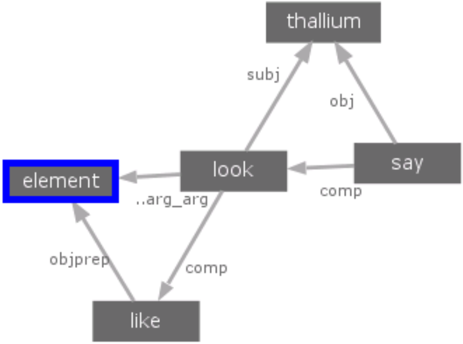


Thallium is said to look like this element.

Syntactic-Semantic Graphs

- Slot Grammar (grammatical relations)
  `subj, obj, iobj, ndet,...`
- Shallow Semantic Relations
  `instanceOf, theme, experiencer`
- Derivational Morphology
  `nobj, obj` → `dm_obj_arg`
- Deep Semantic Relations
  `actorInOpus, bornWhen, nationalityOf,...`
- Coreference (Anaphora resolution)
- Predicate Argument Structure

[McCord et al., 2011]

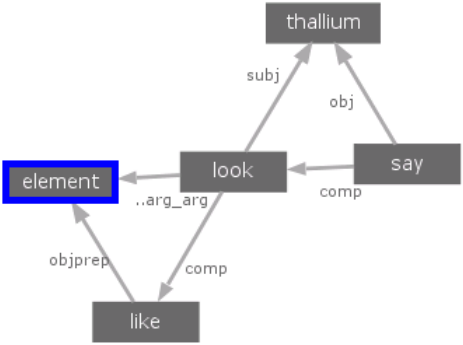


*Thallium is said to look like this **element**.*

Syntactic-Semantic Graphs

- Slot Grammar (grammatical relations)
  `subj, obj, iobj, ndet,...`
- Shallow Semantic Relations
  `instanceOf, theme, experiencer`
- Derivational Morphology
  `nobj, obj` → `dm_obj_arg`
- Deep Semantic Relations
  `actorInOpus, bornWhen, nationalityOf,...`
- Coreference (Anaphora resolution)
- Predicate Argument Structure

[McCord et al., 2011]

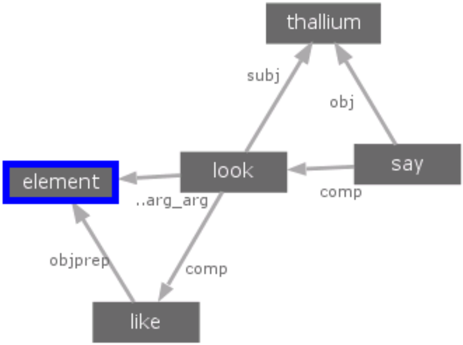


*Thallium is said to look like this **element**.*

Syntactic-Semantic Graphs

- Slot Grammar (grammatical relations)
  `subj, obj, iobj, ndet,...`
- Shallow Semantic Relations
  `instanceOf, theme, experiencer`
- Derivational Morphology
  `nobj, obj` $\rightarrow$ `dm_obj_arg`
- Deep Semantic Relations
  `actorInOpus, bornWhen, nationalityOf,...`
- Coreference (Anaphora resolution)
- Predicate Argument Structure

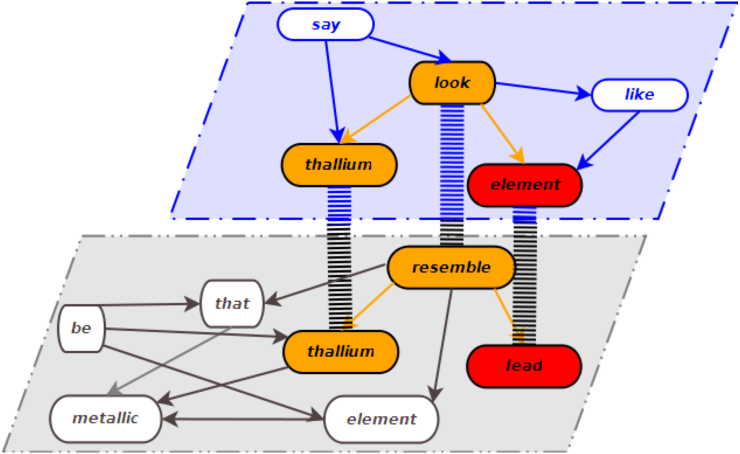
[McCord et al., 2011]



Thallium is said to look like this ***element****.*

LFACS Algorithm

QUESTION: *Thallium is said to look like* ***this element***.



PASSAGE: *Thallium is a metallic element that resembles* ***lead***.

Term Matchers used in LFACS

- **Text Equals**: same lemma
- **Token Overlap**: *Bob Dole* ⇔ *Dole*
- **Wikipedia Redirects**
- **Derivational Morphology**: morphological variants
  destroy ⇔ *destruction*
- **Date / Time Matcher**: *18th century* ⇔ *1754*
- **WordNet Synonyms** for verbs
- **Geospatial Matcher**: *Chicago* ⇔ *United States of America*

⇒ Aggregate Matcher:

returns max of all term matchers

Term Matchers used in LFACS

- **Text Equals**: same lemma
- **Token Overlap**: *Bob Dole* ⇔ *Dole*
- **Wikipedia Redirects**
- **Derivational Morphology**: morphological variants
  destroy ⇔ *destruction*
- **Date / Time Matcher**: *18th century* ⇔ *1754*
- **WordNet Synonyms** for verbs
- **Geospatial Matcher**: *Chicago* ⇔ *United States of America*

⇒ Aggregate Matcher:

returns max of all term matchers

Term Matchers used in LFACS

- **Text Equals**: same lemma
- **Token Overlap**: *Bob Dole* ⇔ *Dole*
- **Wikipedia Redirects**
- **Derivational Morphology**: morphological variants
  destroy ⇔ *destruction*
- **Date / Time Matcher**: *18th century* ⇔ *1754*
- **WordNet Synonyms** for verbs
- **Geospatial Matcher**: *Chicago* ⇔ *United States of America*

⇒ Aggregate Matcher:

returns max of all term matchers

Term Matchers used in LFACS

- **Text Equals**: same lemma
- **Token Overlap**: *Bob Dole* ⇔ *Dole*
- **Wikipedia Redirects**
- **Derivational Morphology**: morphological variants
  destroy ⇔ *destruction*
- **Date / Time Matcher**: *18th century* ⇔ *1754*
- **WordNet Synonyms** for verbs
- **Geospatial Matcher**: *Chicago* ⇔ *United States of America*

⇒ Aggregate Matcher:

returns max of all term matchers

Term Matchers used in LFACS

- **Text Equals**: same lemma
- **Token Overlap**: *Bob Dole* ⇔ *Dole*
- **Wikipedia Redirects**
- **Derivational Morphology**: morphological variants
  destroy ⇔ *destruction*
- **Date / Time Matcher**: *18th century* ⇔ *1754*
- **WordNet Synonyms** for verbs
- **Geospatial Matcher**: *Chicago* ⇔ *United States of America*

⇒ Aggregate Matcher:

returns max of all term matchers

Term Matchers used in LFACS

- **Text Equals**: same lemma
- **Token Overlap**: *Bob Dole* ⇔ *Dole*
- **Wikipedia Redirects**
- **Derivational Morphology**: morphological variants
  destroy ⇔ *destruction*
- **Date / Time Matcher**: *18th century* ⇔ *1754*
- **WordNet Synonyms** for verbs
- **Geospatial Matcher**: *Chicago* ⇔ *United States of America*

⇒ Aggregate Matcher:

returns max of all term matchers

Term Matchers used in LFACS

- **Text Equals**: same lemma
- **Token Overlap**: *Bob Dole* ⇔ *Dole*
- **Wikipedia Redirects**
- **Derivational Morphology**: morphological variants
  destroy ⇔ *destruction*
- **Date / Time Matcher**: *18th century* ⇔ *1754*
- **WordNet Synonyms** for verbs
- **Geospatial Matcher**: *Chicago* ⇔ *United States of America*

⇒ Aggregate Matcher:

returns max of all term matchers

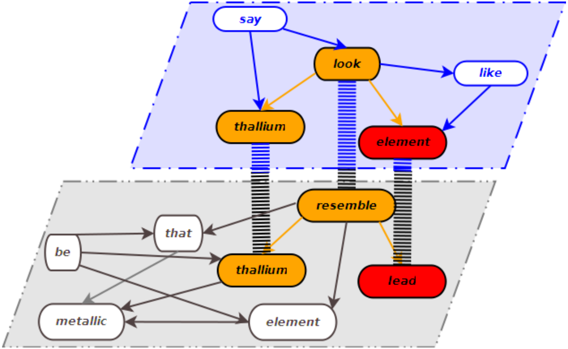
Term Matchers used in LFACS

- **Text Equals**: same lemma
- **Token Overlap**: *Bob Dole* ⇔ *Dole*
- **Wikipedia Redirects**
- **Derivational Morphology**: morphological variants
  destroy ⇔ *destruction*
- **Date / Time Matcher**: *18th century* ⇔ *1754*
- **WordNet Synonyms** for verbs
- **Geospatial Matcher**: *Chicago* ⇔ *United States of America*

⇒ Aggregate Matcher:

returns max of all term matchers

LFACS Algorithm

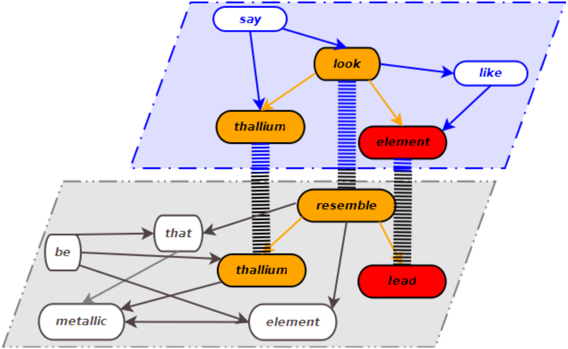


Focus-anchored subgraph (FAS):
set of term match pairs (q_i, p_j) connected to focus / candidate through edge & term matches.
(thallium, thallium)
(like, resemble)
(element, lead)

$$\text{LFACS score} = \sum_{(q_i,p_j)\in FAS} idf(q_i) * matchScore(q_i, p_j)$$

[Murdock et al., 2011]

LFACS Algorithm



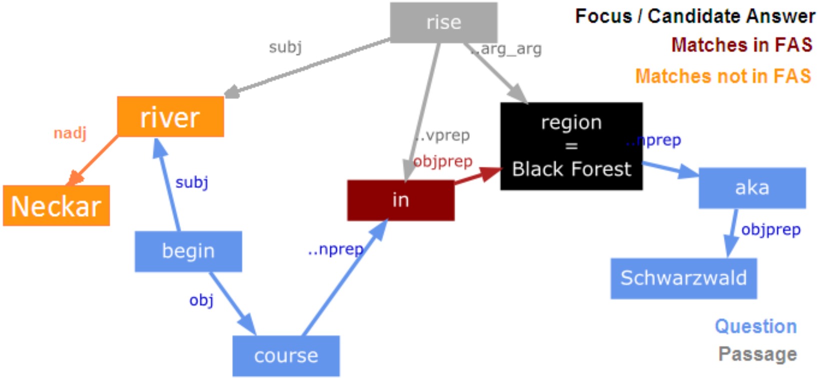
Focus-anchored subgraph (FAS):
set of term match pairs (q_i, p_j) connected to focus / candidate through edge & term matches.
(thallium, thallium)
(like, resemble)
(element, lead)

$$\text{LFACS score} = \sum_{(q_i,p_j)\in FAS} idf(q_i) * matchScore(q_i, p_j)$$

[Murdock et al., 2011]

Outline

Motivation



QUESTION: *The Neckar river begins its course in* ***this region*** *aka the Schwarzwald.*

PASSAGE: *Neckar, river, rising in the* ***Black Forest****.*

Path Scoring Methods

Dependency Path Pair

$$\text{RIVER} \xleftarrow{\textit{subj}} \textit{rise} \xrightarrow{\textit{vprep}} \text{IN}$$

$$\text{RIVER} \xleftarrow{\textit{subj}} \textit{begin} \xrightarrow{\textit{objprep}} \textit{course} \xrightarrow{\textit{nprep}} \text{IN}$$

- Similarity Score for Pair of Dependency Paths:
  - **BSL0.5**: assign 0.5 to any pair of paths
  - **Omiotis**: heuristic based on WordNet semantic relatedness, treats paths as bag-of-words [Tsatsaronis et al., 2009]
  - **Entailment Rules** [Berant et al., 2011]
  - **Dependency Path Similarity Classifier**
- Update LFACS Score:
  for $(q_i, p_j) \in$ matching terms that not in FAS:
  add $\mathit{pathScore}(q_i, p_j) * \mathit{idf}(q_i) * \mathit{matchScore}(q_i, p_j)$

Path Scoring Methods

Dependency Path Pair

$$\text{RIVER} \xleftarrow{subj} rise \xrightarrow{vprep} \text{IN}$$

$$\text{RIVER} \xleftarrow{subj} begin \xrightarrow{objprep} course \xrightarrow{nprep} \text{IN}$$

- Similarity Score for Pair of Dependency Paths:
  - **BSL0.5**: assign 0.5 to any pair of paths
  - **Omiotis**: heuristic based on WordNet semantic relatedness, treats paths as bag-of-words [Tsatsaronis et al., 2009]
  - **Entailment Rules** [Berant et al., 2011]
  - **Dependency Path Similarity Classifier**
- Update LFACS Score:
  for $(q_i, p_j) \in$ matching terms that not in FAS:
  add $pathScore(q_i, p_j) * idf(q_i) * matchScore(q_i, p_j)$

Path Scoring Methods

Dependency Path Pair

$$\text{RIVER} \xleftarrow{subj} rise \xrightarrow{vprep} \text{IN}$$

$$\text{RIVER} \xleftarrow{subj} begin \xrightarrow{objprep} course \xrightarrow{nprep} \text{IN}$$

- Similarity Score for Pair of Dependency Paths:
  - **BSL0.5**: assign 0.5 to any pair of paths
  - **Omiotis**: heuristic based on WordNet semantic relatedness, treats paths as bag-of-words [Tsatsaronis et al., 2009]
  - **Entailment Rules** [Berant et al., 2011]
  - **Dependency Path Similarity Classifier**
- Update LFACS Score:
  for $(q_i, p_j) \in$ matching terms that not in FAS:
  add $pathScore(q_i, p_j) * idf(q_i) * matchScore(q_i, p_j)$

Path Scoring Methods

Dependency Path Pair

$$\text{RIVER} \xleftarrow{\textit{subj}} \textit{rise} \xrightarrow{\textit{vprep}} \text{IN}$$

$$\text{RIVER} \xleftarrow{\textit{subj}} \textit{begin} \xrightarrow{\textit{objprep}} \textit{course} \xrightarrow{\textit{nprep}} \text{IN}$$

- Similarity Score for Pair of Dependency Paths:
  - **BSL0.5**: assign 0.5 to any pair of paths
  - **Omiotis**: heuristic based on WordNet semantic relatedness, treats paths as bag-of-words [Tsatsaronis et al., 2009]
  - **Entailment Rules** [Berant et al., 2011]
  - **Dependency Path Similarity Classifier**
- Update LFACS Score:
  for $(q_i, p_j) \in$ matching terms that not in FAS:
  add $pathScore(q_i, p_j) * idf(q_i) * matchScore(q_i, p_j)$

Path Scoring Methods

Dependency Path Pair

$$\text{RIVER} \xleftarrow{subj} rise \xrightarrow{vprep} \text{IN}$$

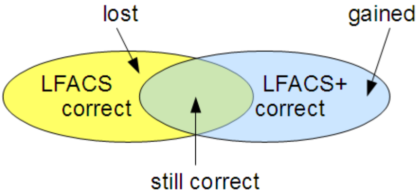
$$\text{RIVER} \xleftarrow{subj} begin \xrightarrow{objprep} course \xrightarrow{nprep} \text{IN}$$

- Similarity Score for Pair of Dependency Paths:
  - **BSL0.5**: assign 0.5 to any pair of paths
  - **Omiotis**: heuristic based on WordNet semantic relatedness, treats paths as bag-of-words [Tsatsaronis et al., 2009]
  - **Entailment Rules** [Berant et al., 2011]
  - **Dependency Path Similarity Classifier**
- Update LFACS Score:
  for $(q_i, p_j) \in$ matching terms that not in FAS:
  add $pathScore(q_i, p_j) * idf(q_i) * matchScore(q_i, p_j)$

Berant's Entailment Rules

Example

(drug) be market by (company)
$\rightarrow$ *(company) manufacture (drug)*



- ignored types & stopwords
- Impact of Berant's rules is minor: 15 gained, 11 lost (out of 3,505)
- Lack of coverage, a lot of rules covered by WordNet relatedness matchers.

	Justification			Type Match	
	yes	partially	no	yes	no
Gained Questions	13	2	0	1	14

Dependency Path Similarity Classifier

Dependency Path Pair

$$\text{(a) RIVER} \xleftarrow{subj} rise \xrightarrow{vprep} \text{IN}$$

$$\text{(b) RIVER} \xleftarrow{subj} begin \xrightarrow{objprep} course \xrightarrow{nprep} \text{IN}$$

- Term Matching: WordNet semantic relatedness

		Path (b) Terms		Path (a) Anchors	MAX per term
		begin	*course*	RIVER	
Path (a) Terms	*rise*	**0.531**	0.0	0.0	0.531

		Path (a) Terms	Path (b) Anchors	MAX per term
		rise	RIVER	
Path (b) Terms	*begin*	0.531	0.0	0.531
	course	0.0	0.307	0.307

MAX_MATCH = 0.531 MIN_MATCH = 0.307

Dependency Path Similarity Classifier

Dependency Path Pair

(a) RIVER $\xleftarrow{subj}$ *rise* $\xrightarrow{vprep}$ IN

(b) RIVER $\xleftarrow{subj}$ *begin* $\xrightarrow{objprep}$ *course* $\xrightarrow{nprep}$ IN

- Term Matching: WordNet semantic relatedness

		Path (b) Terms		Path (a) Anchors	MAX per term
		begin	*course*	RIVER	
Path (a) Terms	*rise*	**0.531**	0.0	0.0	0.531

		Path (a) Terms	Path (b) Anchors	MAX per term
		rise	RIVER	
Path (b) Terms	*begin*	0.531	0.0	0.531
	course	0.0	0.307	0.307

MAX_MATCH = 0.531 MIN_MATCH = 0.307

Dependency Path Similarity Classifier

Dependency Path Pair

(a) RIVER $\xleftarrow{subj}$ *rise* $\xrightarrow{vprep}$ IN

(b) RIVER $\xleftarrow{subj}$ *begin* $\xrightarrow{objprep}$ *course* $\xrightarrow{nprep}$ IN

- Term Matching: WordNet semantic relatedness

		Path (b) Terms		Path (a) Anchors	MAX per term
		begin	*course*	RIVER	
Path (a) Terms	*rise*	**0.531**	0.0	0.0	0.531

		Path (a) Terms	Path (b) Anchors	MAX per term
		rise	RIVER	
Path (b) Terms	*begin*	0.531	0.0	0.531
	course	0.0	0.307	0.307

MAX_MATCH = 0.531 MIN_MATCH = 0.307

Dependency Path Similarity Classifier

Dependency Path Pair

(a) RIVER $\xleftarrow{subj}$ *rise (VERB)* $\xrightarrow{vprep}$ IN

(b) RIVER $\xleftarrow{subj}$ *begin (VERB)* $\xrightarrow{objprep}$ *course (NOUN)* $\xrightarrow{nprep}$ IN

- NONMATCHED_POS_NOUN = 1.0
- NONMATCHED_POS_VERB = 0.0
- NONMATCHED_SLOT_SUBJ = 0.0
- ...

More features based on distributional semantics:

- TWREX_NOUNS_ON_PATHS
- TWREX_VERBS_ON_PATHS
- TWREX_FRAME_PATHS

Dependency Path Classifier

Training Data:

- Sentences extracted from Wikipedia for DBPedia relations. [Wang et al., 2011].
- same relation $\rightarrow$ paths between arguments similar.
- Cleaned using heuristics / manually, kept 417 out of 7,000 relations.
  $\Rightarrow$ about 5,000 dependency path pairs.

Example

(a) **Kay** was educated at **the University of Colorado at Boulder**.

(b) **Kay** graduated from **the University of Colorado at Boulder**.

Dependency Path Classifier

Evaluation on dependency paths for relation instances.
Leaving-One-Out cross validation.
Distribution: 50% similar, 50% not similar

Version	Accuracy	P true	R true	F true
A	63.0	64.1	60.8	62.4
B	66.2	70.8	56.4	62.8
C	67.6	71.0	60.6	65.4

(A) MIN_MATCH, MAX_MATCH, MIN_MATCH_WITH_ANCHORS, MAX_MATCH_WITH_ANCHORS

(B) the above plus TWREX_NOUNS_ON_PATHS, TWREX_VERBS_ON_PATHS and TWREX_FRAME_PATHS

(C) the above plus OMIOTIS, as well as the UNMATCHED_SLOT and UNMATCHED_POS features.

Outline

Component Statistics

- **Recall Count**: correct answer available
- **Fired+Exists Correct**: non-zero score and correct answer candidate available.
- **Correct Expected**: composed of
  - Correct Untied = highest score for correct answer
  - Correct Tied = highest score to correct answer & wrong answer(s)
- **Expected Precision** = Correct Expected / Fired+Exists Correct
- **Expected Recall** = Correct Expected / Recall Count

	Fired+Exists Correct	Expected Precision	Expected Recall	Expected F-measure
PassageTermMatch	3009	44.6	44.6	44.6
LFACS	2425	28.7	23.1	25.6
LFACS+BSL0.5	2645	31.9	28.1	**29.9**
LFACS+Class	2597	30.9	26.7	**28.6**
LFACS+Omiotis	2617	31.7	27.5	**29.5**

Component Statistics

- **Recall Count**:
  correct answer available
- **Fired+Exists Correct**:
  non-zero score and correct answer candidate available.
- **Correct Expected**: composed of
  - Correct Untied = highest score for correct answer
  - Correct Tied = highest score to correct answer & wrong answer(s)
- **Expected Precision** = Correct Expected / Fired+Exists Correct
- **Expected Recall** = Correct Expected / Recall Count

	Fired+Exists Correct	Expected Precision	Expected Recall	Expected F-measure
PassageTermMatch	3009	44.6	44.6	44.6
LFACS	2425	28.7	23.1	25.6
LFACS+BSL0.5	2645	31.9	28.1	**29.9**
LFACS+Class	2597	30.9	26.7	**28.6**
LFACS+Omiotis	2617	31.7	27.5	**29.5**

Component Statistics

- **Recall Count**: correct answer available
- **Fired+Exists Correct**: non-zero score and correct answer candidate available.
- **Correct Expected**: composed of
  - Correct Untied = highest score for correct answer
  - Correct Tied = highest score to correct answer & wrong answer(s)
- **Expected Precision** = Correct Expected / Fired+Exists Correct
- **Expected Recall** = Correct Expected / Recall Count

	Fired+Exists Correct	Expected Precision	Expected Recall	Expected F-measure
PassageTermMatch	3009	44.6	44.6	44.6
LFACS	2425	28.7	23.1	25.6
LFACS+BSL0.5	2645	31.9	28.1	**29.9**
LFACS+Class	2597	30.9	26.7	**28.6**
LFACS+Omiotis	2617	31.7	27.5	**29.5**

Component Statistics

- **Recall Count**:
  correct answer available
- **Fired+Exists Correct**:
  non-zero score and correct answer candidate available.
- **Correct Expected**: composed of
  - Correct Untied = highest score for correct answer
  - Correct Tied = highest score to correct answer & wrong answer(s)
- **Expected Precision** = Correct Expected / Fired+Exists Correct
- **Expected Recall** = Correct Expected / Recall Count

	Fired+Exists Correct	Expected Precision	Expected Recall	Expected F-measure
PassageTermMatch	3009	44.6	44.6	44.6
LFACS	2425	28.7	23.1	25.6
LFACS+BSL0.5	2645	31.9	28.1	**29.9**
LFACS+Class	2597	30.9	26.7	**28.6**
LFACS+Omiotis	2617	31.7	27.5	**29.5**

Component Statistics

- **Recall Count**:
  correct answer available
- **Fired+Exists Correct**:
  non-zero score and correct answer candidate available.
- **Correct Expected**: composed of
  - Correct Untied = highest score for correct answer
  - Correct Tied = highest score to correct answer & wrong answer(s)
- **Expected Precision** = Correct Expected / Fired+Exists Correct
- **Expected Recall** = Correct Expected / Recall Count

	Fired+Exists Correct	Expected Precision	Expected Recall	Expected F-measure
PassageTermMatch	3009	44.6	44.6	44.6
LFACS	2425	28.7	23.1	25.6
LFACS+BSL0.5	2645	31.9	28.1	**29.9**
LFACS+Class	2597	30.9	26.7	**28.6**
LFACS+Omiotis	2617	31.7	27.5	**29.5**

Impact on QA System

Passage Scoring Baseline Stats

	ACCURACY	PREC@70
No passage scorers	55.0	69.9
PTM	58.6	72.7
LFACS	57.1	71.7
LFACS+BSL0.5*	59.0	73.8
LFACS+Class	57.8	72.8
LFACS+Omiotis*	58.2	72.9

* difference VS LFACS statistically significant according to McNemar's test with Yates' correction for continuity.

Correlation Coefficient with Answer Correctness

LFACS	17.35
LFACS+BSL0.5	16.95
LFACS+Class	17.54
LFACS+Omiotis	15.27

Qualitative Analysis - Justification

- Component statistics best for PTM / BSL0.5
  $\Rightarrow$ more relaxed matching $\Rightarrow$ more correct answers.

	Justification			# Questions
	yes	partially	no	
LFACS+BSL0.5	72.8	20.4	6.8	206
LFACS+Class	79.1	17.6	3.3	210
LFACS+Omiotis	75.7	18.8	5.5	202

- BSL0.5 seems to 'guess' more than LFACS+Class and LFACS+Omiotis.

Outline

Summary - Experimental Results

Comparison with LFACS.

More relaxed graph matching

- increases Expected F-measure by 3-4.5%
- increases accuracy of QA system (baseline configuration) by 1-2% (when using only one passage scorer).
- correlation coefficient with answer correctness: LFACS+Class > LFACS

Justification: Analysis for gained questions vs. LFACS

- LFACS+BSL0.5 'guesses' more than LFACS+Class and LFACS+Omiotis.

Summary - Experimental Results

Comparison with LFACS.

More relaxed graph matching

- increases Expected F-measure by 3-4.5%
- increases accuracy of QA system (baseline configuration) by 1-2% (when using only one passage scorer).
- correlation coefficient with answer correctness: LFACS+Class $>$ LFACS

Justification: Analysis for gained questions vs. LFACS

- LFACS+BSL0.5 'guesses' more than LFACS+Class and LFACS+Omiotis.

Summary - Contributions

- Enhanced LFACS with soft matching methods.
- Experimented with various path matching methods.
- We *can* find more correct answers.
- We do so by identifying more *justifying* passages.

Summary - Future Work

- Path matching that uses *edge labels*.
- Gather and leverage more paraphrasing corpora.
- Test in other domains (medical).
- Using tree/sequence/graph kernels for scoring subgraphs between matching terms and focus-anchored subgraph.
  Main issue: training data.
- Improve component-level evaluation metrics for passage scoring.

Questions?

THANKS!

... and Thanks to Bill Murdock, Jennifer Chu-Carroll, Wim De Pauw, Manfred Pinkal, Dietrich Klakow, Karen Ingraffea, Aditya Kalyanpur and all the others who invested their time!

References I

Berant, J., Dagan, I., and Goldberger, J. (2011).
Global learning of typed entailment rules.
In *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA*, pages 610–619.

Fellbaum, C. (1998).
WordNet: An Electronic Lexical Database.
Bradford Books.

References II

- Leacock, C., Miller, G. A., and Chodorow, M. (1998).
  Combining local context with wordnet similarity for word sense identification. in christiane fellbaum.
  In Fellbaum, C., editor, *WordNet: A Lexical Reference System and its Application*, chapter Combining local context with WordNet similarity for word sense identification. MIT Press, Cambridge, MA, USA.

- McCord, M. C., Murdock, J. W., and Boguraev, B. (2011).
  Deep parsing in watson.
  IBM Research and Development Journal Special Issue on DeepQA (in preparation).

References III

Murdock, J. W., Fan, J., Lally, A., Shima, H., and Boguraev, B. (2011).
Textual evidence gathering and analysis.
IBM Research and Development Journal Special Issue on DeepQA (in preparation).

Tsatsaronis, G., Varlamis, I., Vazirgiannis, M., and Nørvåg, K. (2009).
Omiotis: A thesaurus-based measure of text relatedness.
In *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases: Part II*, ECML PKDD '09, pages 742–745, Berlin, Heidelberg. Springer-Verlag.

References IV

Wang, C., Fan, J., Kalyanpur, A. A., and Gondek, D. (2011).
Relation extraction with relation topics.
In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Wu, Z. and Palmer, M. (1994).
Verb semantics and lexical selection.
In *Proc. of the 32nd annual meeting on Association for Computational Linguistics*, pages 133–138.

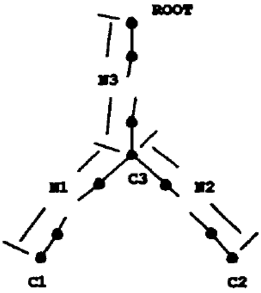
BACKUP SLIDES

WordNet Semantic Relatedness Measures

WORDNET = Thesaurus of English language

Concepts connected by [Fellbaum, 1998]

- hyponomy (is-a)
- antonmy (opposite)
- synonymy
- meronymy (part-of)



$$relatedness_{LCH}(c_1, c_2) = \max\left[log\left(\frac{d(c_1, c_2)}{2D}\right)\right]$$

$$relatedness_{WP}(c_1, c_2) = \frac{2N_3}{N_1 + N_2 + 2N_3}$$

[Wu and Palmer, 1994], [Leacock et al., 1998]

Features for Dependency Path Classifier: Example

(a) CONDUCT $\xrightarrow{vprep}$ *in* $\xrightarrow{objprep}$ *create* $\xrightarrow{obj}$ HISTORY

(b) CONDUCT $\xleftarrow{objprep}$ *by* $\xleftarrow{vprep}$ *obtain* $\xrightarrow{obj}$ *records* $\xrightarrow{nprep}$ *of* $\xrightarrow{objprep}$ HISTORY

		Path (b) Terms		Path (a) Anchors		MAX per term
		obtain	*record*	CONDUCT	HISTORY	
Path (a) Terms	*create*	**0.531**	0.0	0.470	0.0	0.531

		Path (a) Terms	Path (b) Anchors		MAX per term
		create	CONDUCT	HISTORY	
Path (b) Terms	*obtain*	**0.531**	0.421	0.0	0.531
	record	0.0	0.0	0.429	0.429

MIN_MATCH = 0.531

MAX_MATCH = 0.429

Path Scoring Methods: Omiotis

Heuristic that treats dependency paths as bag of words.
[Tsatsaronis et al., 2009]

Example

$$\text{RIVER} \xleftarrow{subj} rise \xrightarrow{vprep} \text{IN}$$

$$\text{RIVER} \xleftarrow{subj} begin \xrightarrow{objprep} course \xrightarrow{nprep} \text{IN}$$

$$Omiotis(P_Q, P_P)$$
$$= \frac{1}{2}\left[\frac{1}{|P_Q|}\sum_{i=1}^{|P_Q|}\left(\lambda_{i,x(i)} * SR(q_i, p_{x(i)})\right) + \frac{1}{|P_P|}\sum_{j=1}^{|P_P|}\left(\lambda_{y(j),j} * SR(q_{y(j)}, p_j)\right)\right]$$

$\lambda_{i,j}$ = harmonic mean of the normalized idfs of q_i and p_j.

$$x(i) = \underset{j\in(1,|P_P|)}{\arg\max}\,[\lambda_{i,j} * SR(q_i, p_j))] \quad y(j) = \underset{i\in(1,|P_Q|)}{\arg\max}\,[\lambda_{i,j} * SR(q_i, p_j))].$$