SI 630

Natural Language Processing: Algorithms and People

Lecture 8 — NLP Tasks: Information Extraction, Textual Entailment, Question Answering

Feb. 26, 2020

Some slides adapted from Chris Manning, Dan Jurafsky, Delip Rao, Yejin Choi, Ido Dagan, Dan Roth, Fabio Massimo Zanzotto, Jonathan Berant, Kai-Wei Chang

Class Updates

Class Updates

- HW2 scores released after spring break

Class Updates

- HW2 scores released after spring break
- Project Proposal revisions due Monday

Class Updates

- HW2 scores released after spring break
- Project Proposal revisions due Monday
    - Can still get full credit

Class Updates

- HW2 scores released after spring break
- Project Proposal revisions due Monday
    - Can still get full credit
    - Project Update deadline is still the same even if you revise

Class Updates

- HW2 scores released after spring break
- Project Proposal revisions due Monday
  - Can still get full credit
  - Project Update deadline is still the same even if you revise
- HW3 is due after Spring Break

Class Updates

- HW2 scores released after spring break
- Project Proposal revisions due Monday
  - Can still get full credit
  - Project Update deadline is still the same even if you revise
- HW3 is due after Spring Break
  - But the GSIs and I will have limited availability over spring break so ask questions now 🙏

Today's Overview

Today's Overview

- What can we do with NLP?

Today's Overview

- What can we do with NLP?
- Four different categories of tasks and algorithms/subtasks within each

Today's Overview

- What can we do with NLP?
- Four different categories of tasks and algorithms/subtasks within each
- More Neural Networks

Meta-Discussion of 630

Meta-Discussion of 630

Meta-Discussion of 630

Fundamentals + Machine Learning

Structure in Language

Semantics + NLP Applications

I'm the NLP Class Train! Choo Choo!

Meta-Discussion of 630

Fundamentals + Machine Learning

Structure in Language

Semantics + NLP Applications

I'm the NLP Class Train! Choo Choo!

Information Extraction

Information Extraction

- Information extraction (IE) systems

Information Extraction

- Information extraction (IE) systems
    - Find and understand limited relevant parts of texts

Information Extraction

- Information extraction (IE) systems
  - Find and understand limited relevant parts of texts
  - Gather information from many pieces of text

Information Extraction

- Information extraction (IE) systems
  - Find and understand limited relevant parts of texts
  - Gather information from many pieces of text
  - Produce a structured representation of relevant information:

Information Extraction

- Information extraction (IE) systems
  - Find and understand limited relevant parts of texts
  - Gather information from many pieces of text
  - Produce a structured representation of relevant information:
    - relations (in the database sense), a.k.a.,
    - a knowledge base

Information Extraction

- Information extraction (IE) systems
  - Find and understand limited relevant parts of texts
  - Gather information from many pieces of text
  - Produce a structured representation of relevant information:
    - relations (in the database sense), a.k.a.,
    - a knowledge base
  - Goals:
    1. Organize information so that it is useful to people

Information Extraction

- Information extraction (IE) systems
  - Find and understand limited relevant parts of texts
  - Gather information from many pieces of text
  - Produce a structured representation of relevant information:
    - relations (in the database sense), a.k.a.,
    - a knowledge base
  - Goals:
    1. Organize information so that it is useful to people
    2. Put information in a semantically precise form that allows further inferences to be made by computer algorithms

Information Extraction (IE)

Information Extraction (IE)

- IE systems extract clear, factual information
  - Roughly: Who did what to whom when?

Information Extraction (IE)

- IE systems extract clear, factual information
  - Roughly: Who did what to whom when?
- For example,

Information Extraction (IE)

- IE systems extract clear, factual information
  - Roughly: Who did what to whom when?
- For example,
  - Gathering earnings, profits, board members, headquarters, etc. from company reports

Information Extraction (IE)

- IE systems extract clear, factual information
  - Roughly: Who did what to whom when?
- For example,
  - Gathering earnings, profits, board members, headquarters, etc. from company reports
    - The headquarters of BHP Billiton Limited, and the global headquarters of the combined BHP Billiton Group, are located in Melbourne, Australia.

Information Extraction (IE)

- IE systems extract clear, factual information
  - Roughly: Who did what to whom when?
- For example,
  - Gathering earnings, profits, board members, headquarters, etc. from company reports
    - The headquarters of BHP Billiton Limited, and the global headquarters of the combined BHP Billiton Group, are located in Melbourne, Australia.
    - headquarters(“BHP Biliton Limited”, “Melbourne, Australia”)

Information Extraction (IE)

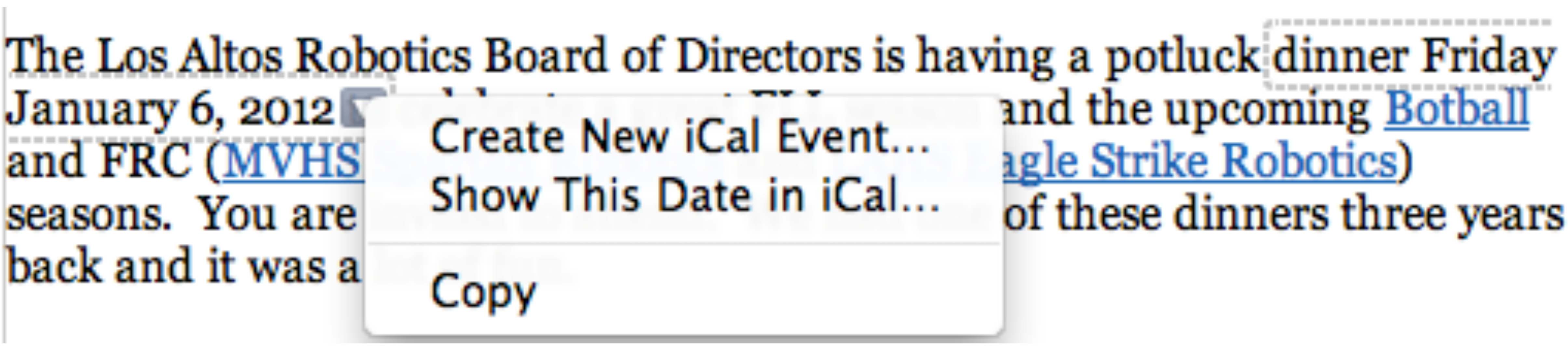
- IE systems extract clear, factual information
  - Roughly: Who did what to whom when?
- For example,
  - Gathering earnings, profits, board members, headquarters, etc. from company reports
    - The headquarters of BHP Billiton Limited, and the global headquarters of the combined BHP Billiton Group, are located in Melbourne, Australia.
    - headquarters(“BHP Biliton Limited”, “Melbourne, Australia”)
  - Learn drug-gene product interactions from medical research literature

Low-level information extraction

- Is now available – and I think popular – in applications like Apple or Google mail, and web indexing

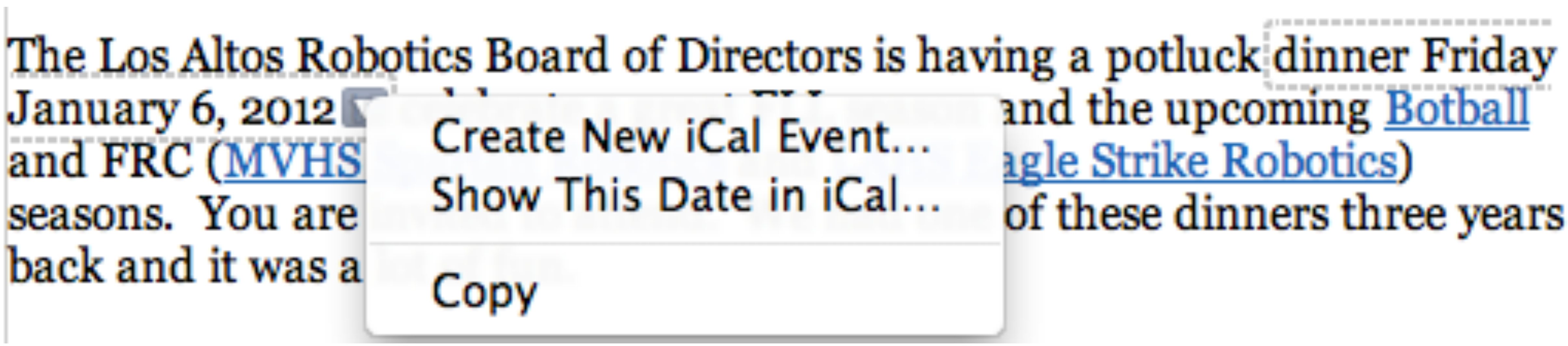
Low-level information extraction

- Is now available – and I think popular – in applications like Apple or Google mail, and web indexing



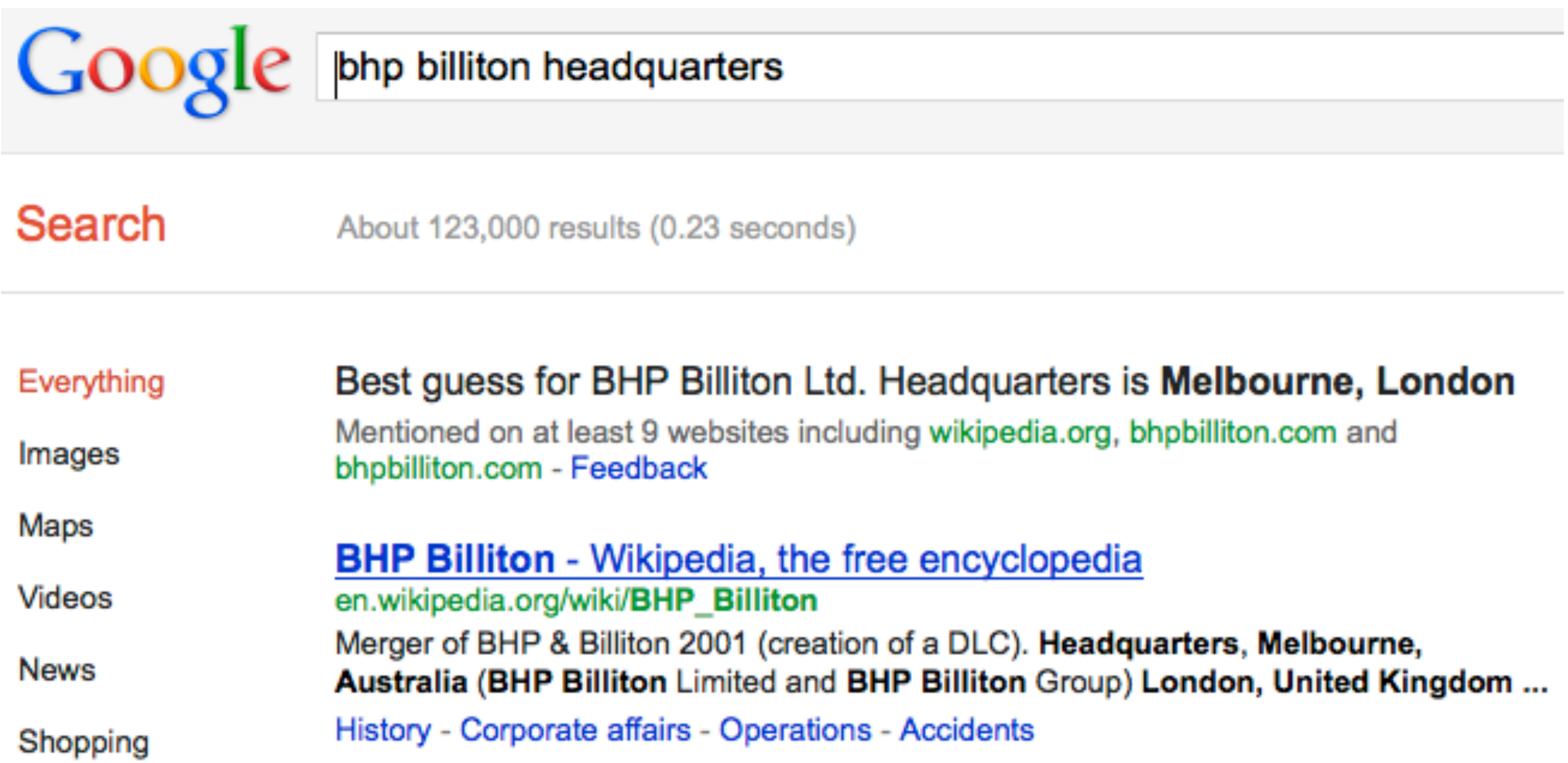
Low-level information extraction

- Is now available – and I think popular – in applications like Apple or Google mail, and web indexing

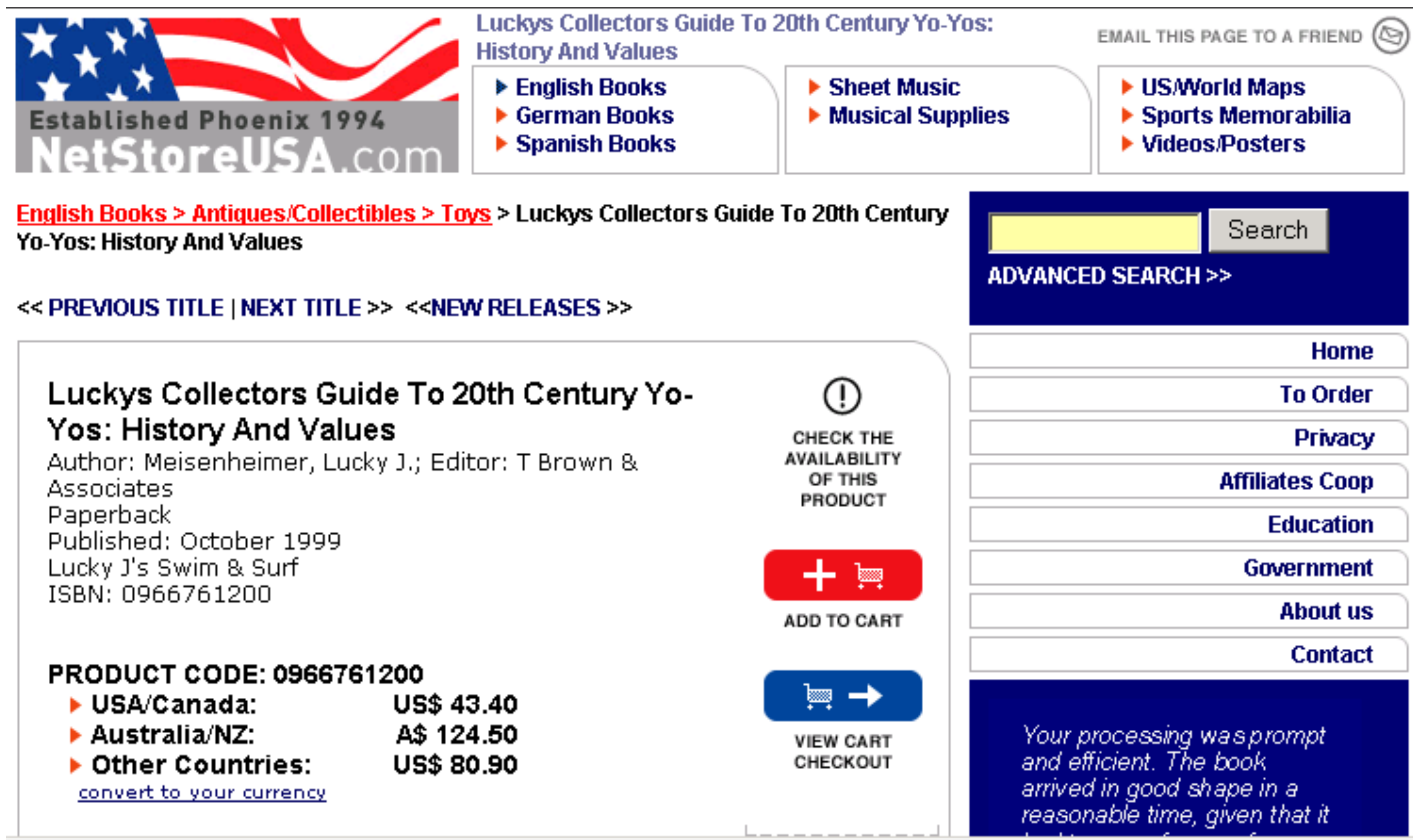


- Often seems to be based on regular expressions and name lists

Low-level information extraction

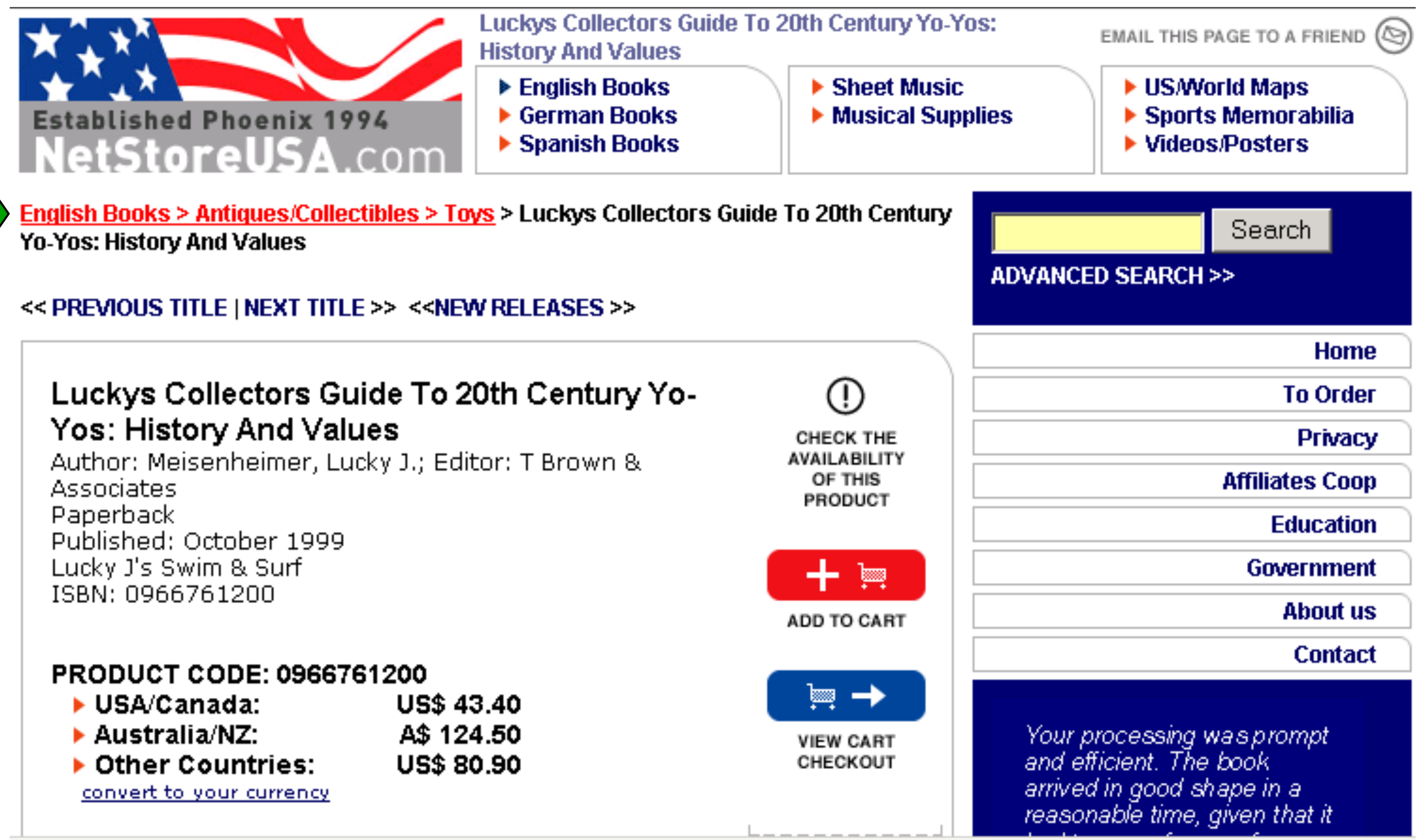


Why is IE hard on the web?



Why is IE hard on the web?

A book,
Not a toy



Why is IE hard on the web?

Why is IE hard on the web?

How is IE useful?

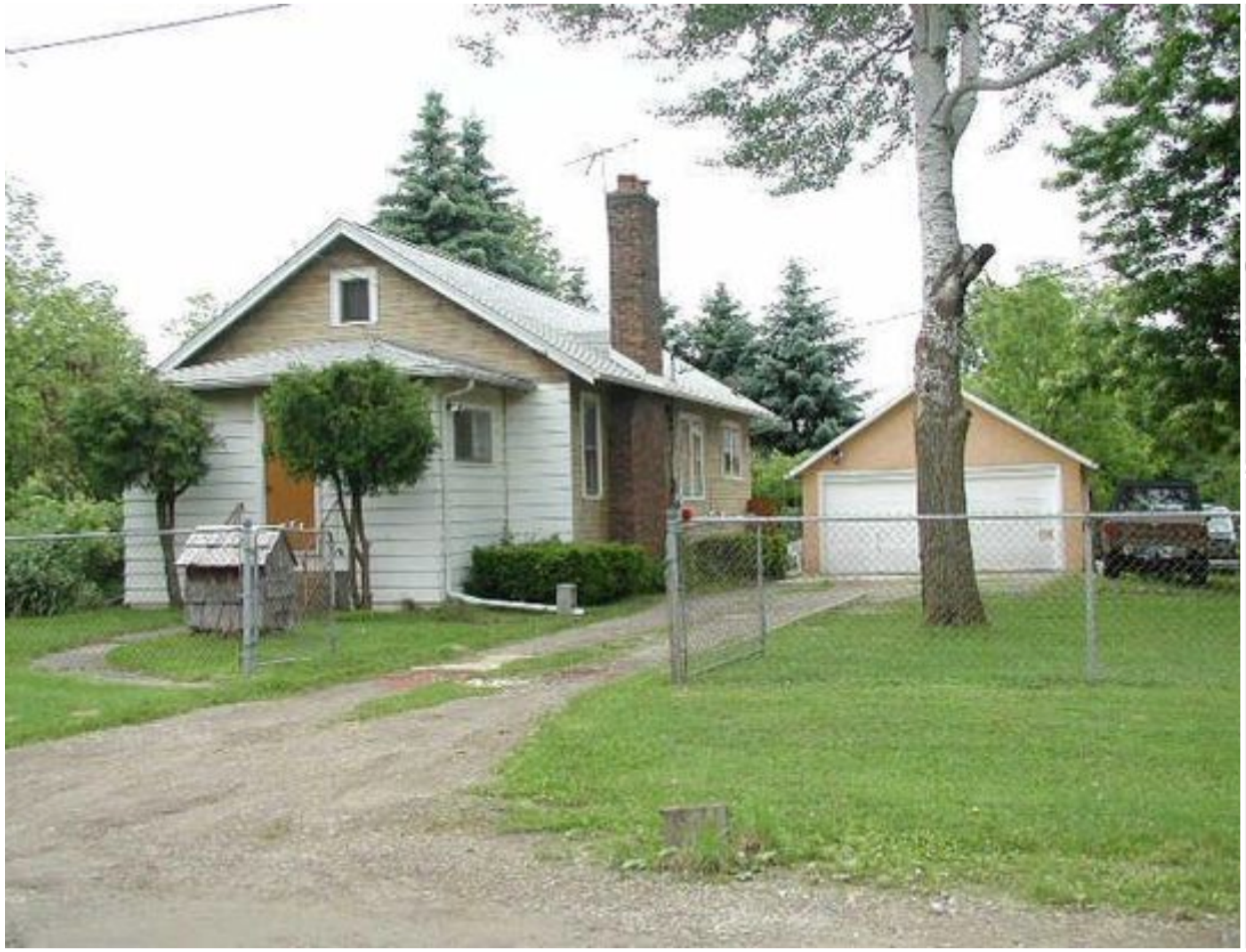
Classified Advertisements (Real Estate)

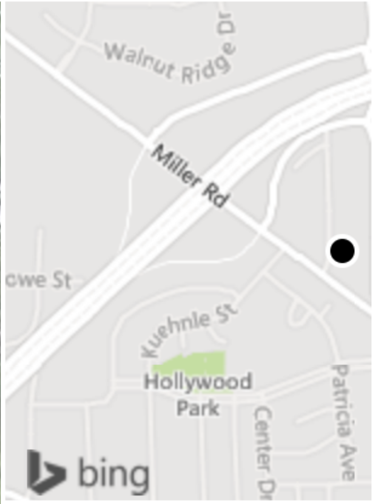
Background:

- Plain text advertisements
- Lowest common denominator: only thing that 70+ newspapers using many different publishing systems can all handle

```
<ADNUM>2067206v1</ADNUM>
<DATE>March 02, 1998</DATE>
<ADTITLE>MADDINGTON $89,000</ADTITLE>
<ADTEXT>
OPEN 1.00 - 1.45<BR>
U 11 / 10 BERTRAM ST<BR>
NEW TO MARKET Beautiful<BR>
3 brm freestanding<BR>
villa, close to shops & bus<BR>
Owner moved to Melbourne<BR>
ideally suit 1st home buyer,<BR>
investor & 55 and over.<BR>
Brian Hazelden 0418 958 996<BR>
R WHITE LEEMING 9332 3477
</ADTEXT>
```

Public View Owner View





1650 Calvin St,
Ann Arbor, MI 48103

2 beds · 1 bath · 962 sqft

FOR SALE
$150,000
Zestimate®: $159,728

EST. MORTGAGE
$604/mo
Get pre-qualified

This 962 square foot single family home has 2 bedrooms and 1.0 bathrooms. It is located at 1650 Calvin St Ann Arbor, Michigan.

Facts and Features

Type
Single Family

Year Built
1929

Heating
No Data

Why doesn't text search (IR) work?

Why doesn't text search (IR) work?

What you search for in real estate advertisements:

Why doesn't text search (IR) work?

What you search for in real estate advertisements:

- Town/suburb. You might think easy, but:
  - Real estate agents: Coldwell Banker, Mosman
  - Phrases: Only 45 minutes from Ypsi
  - Multiple property ads have different suburbs in one ad

Why doesn't text search (IR) work?

What you search for in real estate advertisements:

- Town/suburb. You might think easy, but:
  - Real estate agents: Coldwell Banker, Mosman
  - Phrases: Only 45 minutes from Ypsi
  - Multiple property ads have different suburbs in one ad

- Money: want a range not a textual match
  - Multiple amounts: was $155K, now $145K
  - Variations: offers in the high 700s [*but not* rents for $270]

Why doesn't text search (IR) work?

What you search for in real estate advertisements:

- Town/suburb. You might think easy, but:
  - Real estate agents: Coldwell Banker, Mosman
  - Phrases: Only 45 minutes from Ypsi
  - Multiple property ads have different suburbs in one ad
- Money: want a range not a textual match
  - Multiple amounts: was $155K, now $145K
  - Variations: offers in the high 700s [*but not* rents for $270]
- Bedrooms: similar issues: br, bdr, beds, B/R

star
INSIDE MICHELLE HEATON'S STAR-STUDDED BABY SHOWER AT HER HOME
EXCLUSIVE INTERVIEW AND PICTURES
WHY AMY IS SLAGGING OFF JORDAN... AND PETE
"Lee's behaviour is socially unacceptable!"
Most deluded TV love rat ever?
OK!
FIRST FOR CELEBRITY
ISSUE 914 • JAN 28 2014
CELEBRITY BIG BROTHER'S
JASMINE
EXCLUSIVE INTERVIEW AND PICTURES
'MY FEELINGS FOR LEE ARE
This week's hottest celebrity news
heat
KIM & KAN
Lisa's
HOCK
weight
gain
OUT

Relation Extraction

Extracting relations from text

- Company report: “International Business Machines Corporation (IBM or the company) was incorporated in the State of New York on June 16, 1911, as the Computing-Tabulating-Recording Co. (C-T-R)…”

Extracting relations from text

- Company report: "International Business Machines Corporation (IBM or the company) was incorporated in the State of New York on June 16, 1911, as the Computing-Tabulating-Recording Co. (C-T-R)…"

- Extracted Complex Relation:

Company-Founding

Company	IBM
Location	New York
Date	June 16, 1911
Original-Name	Computing-Tabulating-Recording Co.

Extracting relations from text

- Company report: "International Business Machines Corporation (IBM or the company) was incorporated in the State of New York on June 16, 1911, as the Computing-Tabulating-Recording Co. (C-T-R)…"

- Extracted Complex Relation:

  Company-Founding
  - Company IBM
  - Location New York
  - Date June 16, 1911
  - Original-Name Computing-Tabulating-Recording Co.

- But we will focus on the simpler task of extracting relation **triples**

  Founding-year(IBM,1911)

  Founding-location(IBM,New York)

Traditional Relation Extraction

- Approaches:
  - Supervised (ACE)
  - Rule-Based (Hearst Patterns, etc…)

Traditional Relation Extraction

- Approaches:
  - Supervised (ACE)
  - Rule-Based (Hearst Patterns, etc…)
- Advantages:
  - Good in-domain performance
- Disadvantages:
  - Requires lots of human effort for each new relation…
  - Doesn’t generalize much beyond the domain

Hearst Patterns for learning Hypernymy (an X is a Y)

- NP such as {NP,}* {(or | and} NP
- such NP as {NP,}* {(or | and)} NP
- NP {, NP}* (,} (or | and) other NP
- NP {,} (including | especially) {NP,}* (or | and) NP …

Hearst Patterns for learning Hypernymy (an X is a Y)

- NP such as {NP,}* {(or | and} NP
- such NP as {NP,}* {(or | and)} NP
- NP {, NP}* (,} (or | and) other NP
- NP {,} (including | especially) {NP,}* (or | and) NP …
- PRO: Designed for very high precision.

Hearst Patterns for learning Hypernymy (an X is a Y)

- NP such as {NP,}* {(or | and} NP
- such NP as {NP,}* {(or | and)} NP
- NP {, NP}* (,} (or | and) other NP
- NP {,} (including | especially) {NP,}* (or | and) NP …
- PRO: Designed for very high precision.
- CONs: But low recall. Only cover "is-a", later, extended to "part-of" relation – more like "typing". Unclear if such patterns can be signed for all relations/attributes.

Bootstrapping

- <Mark Twain, Elmira> Seed tuple
  - Grep (google) for the environments of the seed tuple

  "Mark Twain is buried in Elmira, NY."

    X is buried in Y

  "The grave of Mark Twain is in Elmira"

    The grave of X is in Y

  "Elmira is Mark Twain's final resting place"

    Y is X's final resting place.

- Use those patterns to grep for new tuples
- Iterate

Dipre: Extract <author,book> pairs

Author	Book
Isaac Asimov	The Robots of Dawn
David Brin	Startide Rising
James Gleick	Chaos: Making a New Science
Charles Dickens	Great Expectations
William Shakespeare	The Comedy of Errors

Brin, Sergei. 1998. Extracting Patterns and Relations from the World Wide Web.

Dipre: Extract <author,book> pairs

- Start with 5 seeds:

Author	Book
Isaac Asimov	The Robots of Dawn
David Brin	Startide Rising
James Gleick	Chaos: Making a New Science
Charles Dickens	Great Expectations
William Shakespeare	The Comedy of Errors

Brin, Sergei. 1998. Extracting Patterns and Relations from the World Wide Web.

Dipre: Extract <author,book> pairs

- Start with 5 seeds:

Author	Book
Isaac Asimov	The Robots of Dawn
David Brin	Startide Rising
James Gleick	Chaos: Making a New Science
Charles Dickens	Great Expectations
William Shakespeare	The Comedy of Errors

- Find Instances:

The Comedy of Errors, by William Shakespeare, was
The Comedy of Errors, by William Shakespeare, is
The Comedy of Errors, one of William Shakespeare's earliest attempts
The Comedy of Errors, one of William Shakespeare's most

Brin, Sergei. 1998. Extracting Patterns and Relations from the World Wide Web.

Dipre: Extract <author,book> pairs

- Start with 5 seeds:

Author	Book
Isaac Asimov	The Robots of Dawn
David Brin	Startide Rising
James Gleick	Chaos: Making a New Science
Charles Dickens	Great Expectations
William Shakespeare	The Comedy of Errors

- Find Instances:

  The Comedy of Errors, by William Shakespeare, was
  The Comedy of Errors, by William Shakespeare, is
  The Comedy of Errors, one of William Shakespeare's earliest attempts
  The Comedy of Errors, one of William Shakespeare's most

- Extract patterns (group by middle, take longest common prefix/suffix)

  `?x , by ?y ,` `?x , one of ?y 's`

Brin, Sergei. 1998. Extracting Patterns and Relations from the World Wide Web.

Dipre: Extract <author,book> pairs

- Start with 5 seeds:

Author	Book
Isaac Asimov	The Robots of Dawn
David Brin	Startide Rising
James Gleick	Chaos: Making a New Science
Charles Dickens	Great Expectations
William Shakespeare	The Comedy of Errors

- Find Instances:

  The Comedy of Errors, by William Shakespeare, was
  The Comedy of Errors, by William Shakespeare, is
  The Comedy of Errors, one of William Shakespeare's earliest attempts
  The Comedy of Errors, one of William Shakespeare's most

- Extract patterns (group by middle, take longest common prefix/suffix)

  `?x , by ?y ,` `?x , one of ?y 's`

- Now iterate, finding new seeds that match the pattern

Brin, Sergei. 1998. Extracting Patterns and Relations from the World Wide Web.

Distant Supervision

Snow, Jurafsky, Ng. 2005. Learning syntactic patterns for automatic hypernym discovery. NIPS 05
Fei Wu and Daniel S. Weld. 2007. Autonomously Semantifying Wikipeida. CIKM 2007
Mintz, Bills, Snow, Jurafsky. 2009. Distant supervision for relation extraction without labeled data. ACL09

Distant Supervision

- Combine bootstrapping with supervised learning
  - Instead of 5 seeds,
    - Use a large database to get huge # of seed examples
  - Create lots of features from all these examples
  - Combine in a supervised classifier

Snow, Jurafsky, Ng. 2005. Learning syntactic patterns for automatic hypernym discovery. NIPS 05
Fei Wu and Daniel S. Weld. 2007. Autonomously Semantifying Wikipeida. CIKM 2007
Mintz, Bills, Snow, Jurafsky. 2009. Distant supervision for relation extraction without labeled data. ACL09

Distant supervision paradigm

- Like supervised classification:
  - Uses a classifier with lots of features
  - Supervised by detailed hand-created knowledge
  - Doesn't require iteratively expanding patterns

Distant supervision paradigm

- Like supervised classification:
  - Uses a classifier with lots of features
  - Supervised by detailed hand-created knowledge
  - Doesn't require iteratively expanding patterns
- Like unsupervised classification:
  - Uses very large amounts of unlabeled data
  - Not sensitive to genre issues in training corpus

Distantly supervised learning of relation extraction patterns

Distantly supervised learning of relation extraction patterns

1. For each relation

Born-In

Distantly supervised learning of relation extraction patterns

1. For each relation

Born-In

2. For each tuple in big database

<Edwin Hubble, Marshfield>
<Albert Einstein, Ulm>

Distantly supervised learning of relation extraction patterns

1. For each relation

2. For each tuple in big database

3. Find sentences in large corpus with both entities

Born-In

<Edwin Hubble, Marshfield>
<Albert Einstein, Ulm>

Hubble was born in Marshfield

Einstein, born (1879), Ulm

Hubble’s birthplace in Marshfield

Distantly supervised learning of relation extraction patterns

1. For each relation

   Born-In

2. For each tuple in big database

   <Edwin Hubble, Marshfield>
   <Albert Einstein, Ulm>

3. Find sentences in large corpus with both entities

   Hubble was born in Marshfield
   Einstein, born (1879), Ulm
   Hubble's birthplace in Marshfield

4. Extract frequent features (parse, words, etc)

   PER was born in LOC
   PER, born (XXXX), LOC
   PER's birthplace in LOC

Distantly supervised learning of relation extraction patterns

1. For each relation	Born-In
2. For each tuple in big database	<Edwin Hubble, Marshfield> <Albert Einstein, Ulm>
3. Find sentences in large corpus with both entities	Hubble was born in Marshfield Einstein, born (1879), Ulm Hubble's birthplace in Marshfield
4. Extract frequent features (parse, words, etc)	PER was born in LOC PER, born (XXXX), LOC PER's birthplace in LOC
5. Train supervised classifier using thousands of patterns	$P(\text{born-in} \mid f_1, f_2, f_3, \ldots, f_{70000})$

Relation name	Size	Example
/people/person/nationality	281,107	John Dugard, South Africa
/location/location/contains	253,223	Belgium, Nijlen
/people/person/profession	208,888	Dusa McDuff, Mathematician
/people/person/place_of_birth	105,799	Edwin Hubble, Marshfield
/dining/restaurant/cuisine	86,213	MacAyo's Mexican Kitchen, Mexican
/business/business_chain/location	66,529	Apple Inc., Apple Inc., South Park, NC
/biology/organism_classification_rank	42,806	Scorpaeniformes, Order
/film/film/genre	40,658	Where the Sidewalk Ends, Film noir
/film/film/language	31,103	Enter the Phoenix, Cantonese
/biology/organism_higher_classification	30,052	Calopteryx, Calopterygidae
/film/film/country	27,217	Turtle Diary, United States
/film/writer/film	23,856	Irving Shulman, Rebel Without a Cause
/film/director/film	23,539	Michael Mann, Collateral
/film/producer/film	22,079	Diane Eskenazi, Aladdin
/people/deceased_person/place_of_death	18,814	John W. Kern, Asheville
/music/artist/origin	18,619	The Octopus Project, Austin
/people/person/religion	17,582	Joseph Chartrand, Catholicism
/book/author/works_written	17,278	Paul Auster, Travels in the Scriptorium
/soccer/football_position/players	17,244	Midfielder, Chen Tao
/people/deceased_person/cause_of_death	16,709	Richard Daintree, Tuberculosis
/book/book/genre	16,431	Pony Soldiers, Science fiction
/film/film/music	14,070	Stavisky, Stephen Sondheim
/business/company/industry	13,805	ATS Medical, Health care

Table 2: The 23 largest Freebase relations we use, with their size and an instance of each relation.

Feature type	Left window	NE1	Middle	NE2	Right window
Lexical	[]	PER	[was/VERB born/VERB in/CLOSED]	LOC	[]
Lexical	[Astronomer]	PER	[was/VERB born/VERB in/CLOSED]	LOC	[,]
Lexical	[#PAD#, Astronomer]	PER	[was/VERB born/VERB in/CLOSED]	LOC	[, Missouri]
Syntactic	[]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[]
Syntactic	[Edwin Hubble $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[]
Syntactic	[Astronomer $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[]
Syntactic	[]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{lex-mod}$,]
Syntactic	[Edwin Hubble $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{lex-mod}$,]
Syntactic	[Astronomer $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{lex-mod}$,]
Syntactic	[]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{inside}$ Missouri]
Syntactic	[Edwin Hubble $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{inside}$ Missouri]
Syntactic	[Astronomer $\Downarrow_{lex-mod}$]	PER	[$\Uparrow_s$ was $\Downarrow_{pred}$ born $\Downarrow_{mod}$ in $\Downarrow_{pcomp-n}$]	LOC	[$\Downarrow_{inside}$ Missouri]

Table 3: Features for ‘Astronomer Edwin Hubble was born in Marshfield, Missouri’.

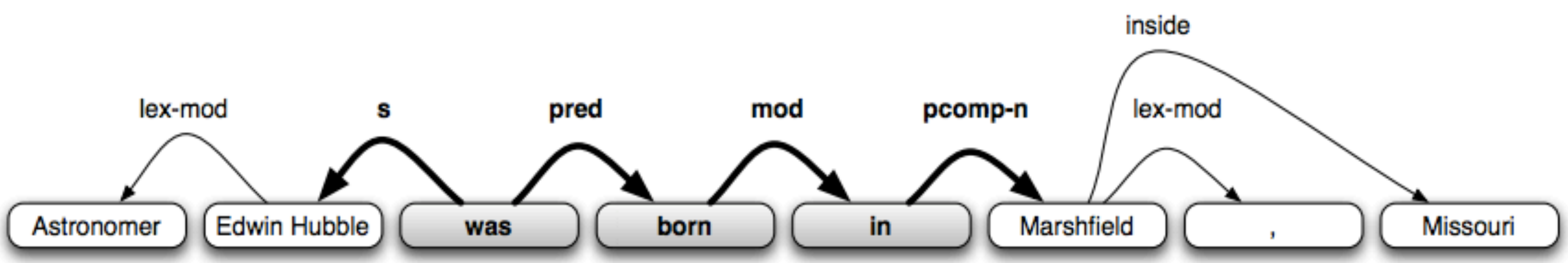


Figure 1: Dependency parse with dependency path from ‘Edwin Hubble’ to ‘Marshfield’ highlighted in boldface.

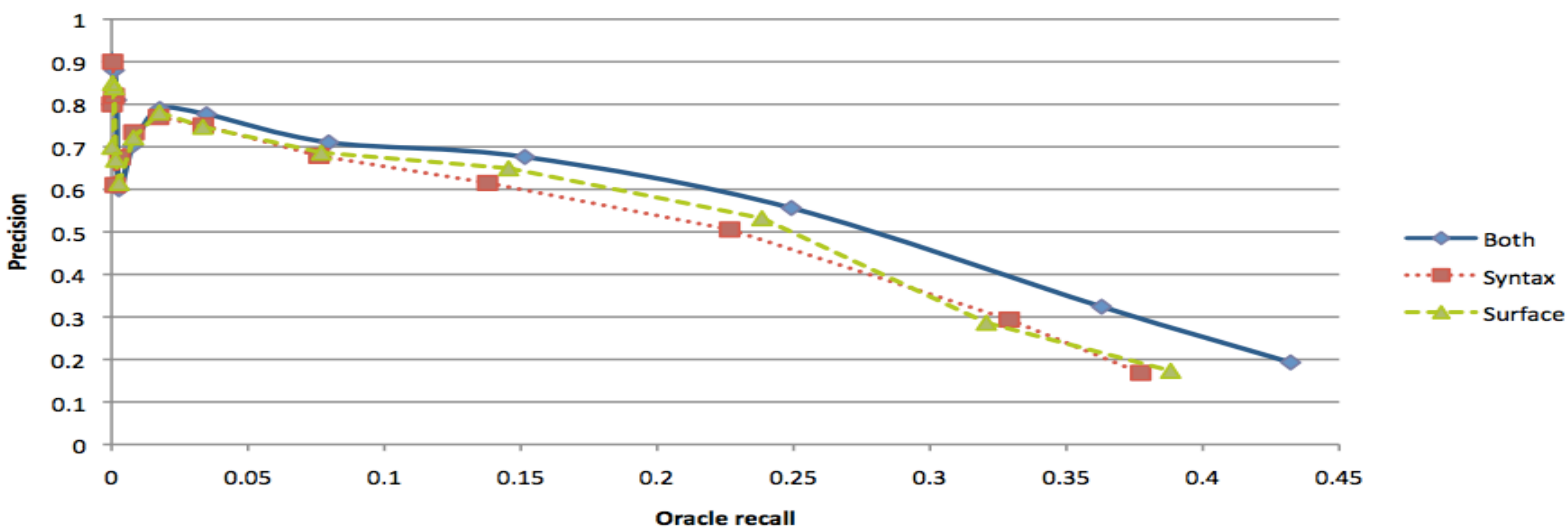
Relation	Feature type	Left window	NE1	Middle	NE2	Right window
/architecture/structure/architect	LEX$\curvearrowleft$		ORG	, the designer of the	PER	
	SYN	designed $\Uparrow_s$	ORG	$\Uparrow_s$ designed $\Downarrow_{by-subj}$ by $\Downarrow_{pcn}$	PER	$\Uparrow_s$ designed
/book/author/works_written	LEX		PER	s novel	ORG	
	SYN		PER	$\Uparrow_{pcn}$ by $\Uparrow_{mod}$ story $\Uparrow_{pred}$ is $\Downarrow_s$	ORG	
/book/book_edition/author_editor	LEX$\curvearrowleft$		ORG	s novel	PER	
	SYN		PER	$\Uparrow_{nn}$ series $\Downarrow_{gen}$	PER	
/business/company/founders	LEX		ORG	co - founder	PER	
	SYN		ORG	$\Uparrow_{nn}$ owner $\Downarrow_{person}$	PER	
/business/company/place_founded	LEX$\curvearrowleft$		ORG	- based	LOC	
	SYN		ORG	$\Uparrow_s$ founded $\Downarrow_{mod}$ in $\Downarrow_{pcn}$	LOC	
/film/film/country	LEX		PER	, released in	LOC	
	SYN	opened $\Uparrow_s$	ORG	$\Uparrow_s$ opened $\Downarrow_{mod}$ in $\Downarrow_{pcn}$	LOC	$\Uparrow_s$ opened
/geography/river/mouth	LEX		LOC	, which flows into the	LOC	
	SYN	the $\Downarrow_{det}$	LOC	$\Uparrow_s$ is $\Downarrow_{pred}$ tributary $\Downarrow_{mod}$ of $\Downarrow_{pcn}$	LOC	$\Downarrow_{det}$ the
/government/political_party/country	LEX$\curvearrowleft$		ORG	politician of the	LOC	
	SYN	candidate $\Uparrow_{nn}$	ORG	$\Uparrow_{nn}$ candidate $\Downarrow_{mod}$ for $\Downarrow_{pcn}$	LOC	$\Uparrow_{nn}$ candidate
/influence/influence_node/influenced	LEX$\curvearrowleft$		PER	, a student of	PER	
	SYN	of $\Uparrow_{pcn}$	PER	$\Uparrow_{pcn}$ of $\Uparrow_{mod}$ student $\Uparrow_{appo}$	PER	$\Uparrow_{pcn}$ of
/language/human_language/region	LEX		LOC	- speaking areas of	LOC	
	SYN		LOC	$\Uparrow_{lex-mod}$ speaking areas $\Downarrow_{mod}$ of $\Downarrow_{pcn}$	LOC	
/music/artist/origin	LEX$\curvearrowleft$		ORG	based band	LOC	
	SYN	is $\Uparrow_s$	ORG	$\Uparrow_s$ is $\Downarrow_{pred}$ band $\Downarrow_{mod}$ from $\Downarrow_{pcn}$	LOC	$\Uparrow_s$ is
/people/deceased_person/place_of_death	LEX		PER	died in	LOC	
	SYN	hanged $\Uparrow_s$	PER	$\Uparrow_s$ hanged $\Downarrow_{mod}$ in $\Downarrow_{pcn}$	LOC	$\Uparrow_s$ hanged
/people/person/nationality	LEX		PER	is a citizen of	LOC	
	SYN		PER	$\Downarrow_{mod}$ from $\Downarrow_{pcn}$	LOC	
/people/person/parents	LEX		PER	, son of	PER	
	SYN	father $\Uparrow_{gen}$	PER	$\Uparrow_{gen}$ father $\Downarrow_{person}$	PER	$\Uparrow_{gen}$ father
/people/person/place_of_birth	LEX$\curvearrowleft$		PER	is the birthplace of	PER	
	SYN		PER	$\Uparrow_s$ born $\Downarrow_{mod}$ in $\Downarrow_{pcn}$	LOC	
/people/person/religion	LEX		PER	embraced	LOC	
	SYN	convert $\Downarrow_{appo}$	PER	$\Downarrow_{appo}$ convert $\Downarrow_{mod}$ to $\Downarrow_{pcn}$	LOC	$\Downarrow_{appo}$ convert

Table 4: Examples of high-weight features for several relations. Key: SYN = syntactic feature; LEX = lexical feature; $\curvearrowleft$ = reversed; NE# = named entity tag of entity.

Automatic Evaluation

- Hold out facts from freebase
  - Evaluate precision and recall
- Problems:
  - Extractions often missing from Freebase
  - Marked as precision errors
  - These are the extractions we really care about! New facts, not contained in Freebase

Automatic Evaluation



Automatic Evaluation: Discussion

- Correct predictions will be missing form DB
  - Underestimates precision

Automatic Evaluation: Discussion

- Correct predictions will be missing form DB
    - Underestimates precision
- This evaluation is biased
    - Systems which make predictions for more frequent entity pairs will do better.
    - Hard constraints => explicitly trained to predict facts already in Freebase

[Riedel et. al. 2013]

Human Evaluation

Relation name	100 instances			1000 instances		
	Syn	Lex	Both	Syn	Lex	Both
/film/director/film	**0.49**	0.43	0.44	**0.49**	0.41	0.46
/film/writer/film	**0.70**	0.60	0.65	**0.71**	0.61	0.69
/geography/river/basin_countries	0.65	0.64	**0.67**	**0.73**	0.71	0.64
/location/country/administrative_divisions	0.68	0.59	**0.70**	**0.72**	0.68	**0.72**
/location/location/contains	0.81	**0.89**	0.84	**0.85**	0.83	0.84
/location/us_county/county_seat	0.51	0.51	**0.53**	0.47	**0.57**	0.42
/music/artist/origin	0.64	0.66	**0.71**	0.61	**0.63**	0.60
/people/deceased_person/place_of_death	0.80	0.79	**0.81**	0.80	**0.81**	0.78
/people/person/nationality	0.61	0.70	**0.72**	0.56	0.61	**0.63**
/people/person/place_of_birth	**0.78**	0.77	**0.78**	0.88	0.85	**0.91**
Average	0.67	0.66	**0.69**	**0.68**	0.67	0.67

Open Information Extraction

Relation Extraction Bottlenecks

Relation Extraction Bottlenecks

- Traditional relation extraction assumes a relation vocabulary.
  - Need to anticipate the knowledge needs in advance.

Relation Extraction Bottlenecks

- Traditional relation extraction assumes a relation vocabulary.
    - Need to anticipate the knowledge needs in advance.
- Typically need a few seed examples or instances per relation.
    - Distant supervision mitigates this somewhat but still assumes the relations are given via a database.

Relation Extraction Bottlenecks

- Traditional relation extraction assumes a relation vocabulary.
  - Need to anticipate the knowledge needs in advance.
- Typically need a few seed examples or instances per relation.
  - Distant supervision mitigates this somewhat but still assumes the relations are given via a database.
- Doesn’t easily scale to large collections such as the web.
  - Need to run each relation classifier on each sentence!

Open Information Extraction

Open Information Extraction

- Identify relation phrases directly from text.

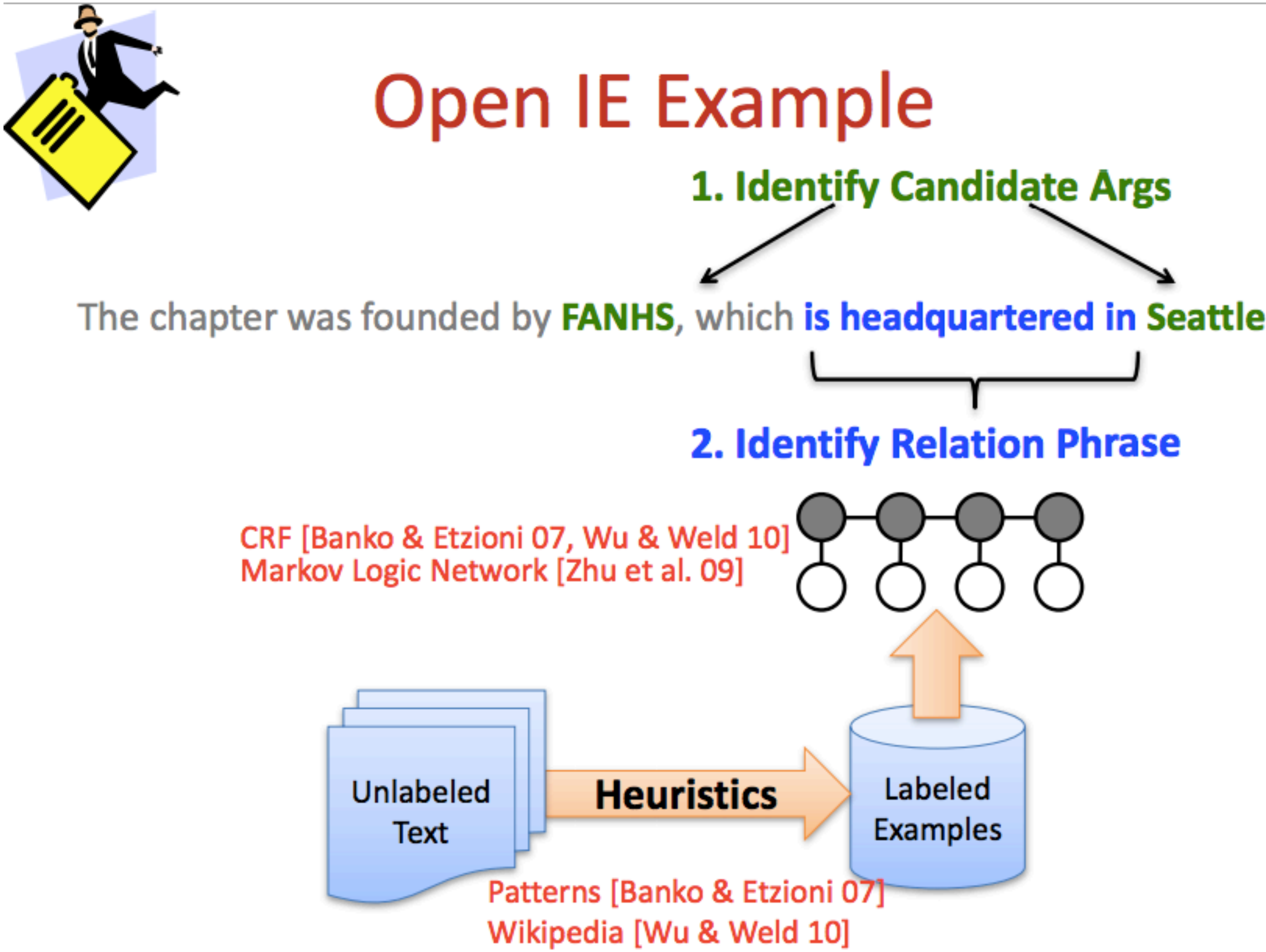
Open Information Extraction

- Identify relation phrases directly from text.
- Avoids lexical patterns.
  - Extractors are specified via POS tags and closed-class words.

Open Information Extraction

- Identify relation phrases directly from text.
- Avoids lexical patterns.
  - Extractors are specified via POS tags and closed-class words.
- Focus on generic ways in which relations are expressed.
  - Not domain specific.

Text Runner



Open IE Example
1. Identify Candidate Args
The chapter was founded by FANHS, which is headquartered in Seattle.
2. Identify Relation Phrase
CRF [Banko & Etzioni 07, Wu & Weld 10]
Markov Logic Network [Zhu et al. 09]
Unlabeled Text
Heuristics
Labeled Examples
Patterns [Banko & Etzioni 07]
Wikipedia [Wu & Weld 10]

Two Issues

Incoherent Extractions

The guide contains dead links and omits sites.

(**The guide**, **contains omits**, **sites**)

Extendicare agreed to buy Arbor for about $432M in assumed debt.

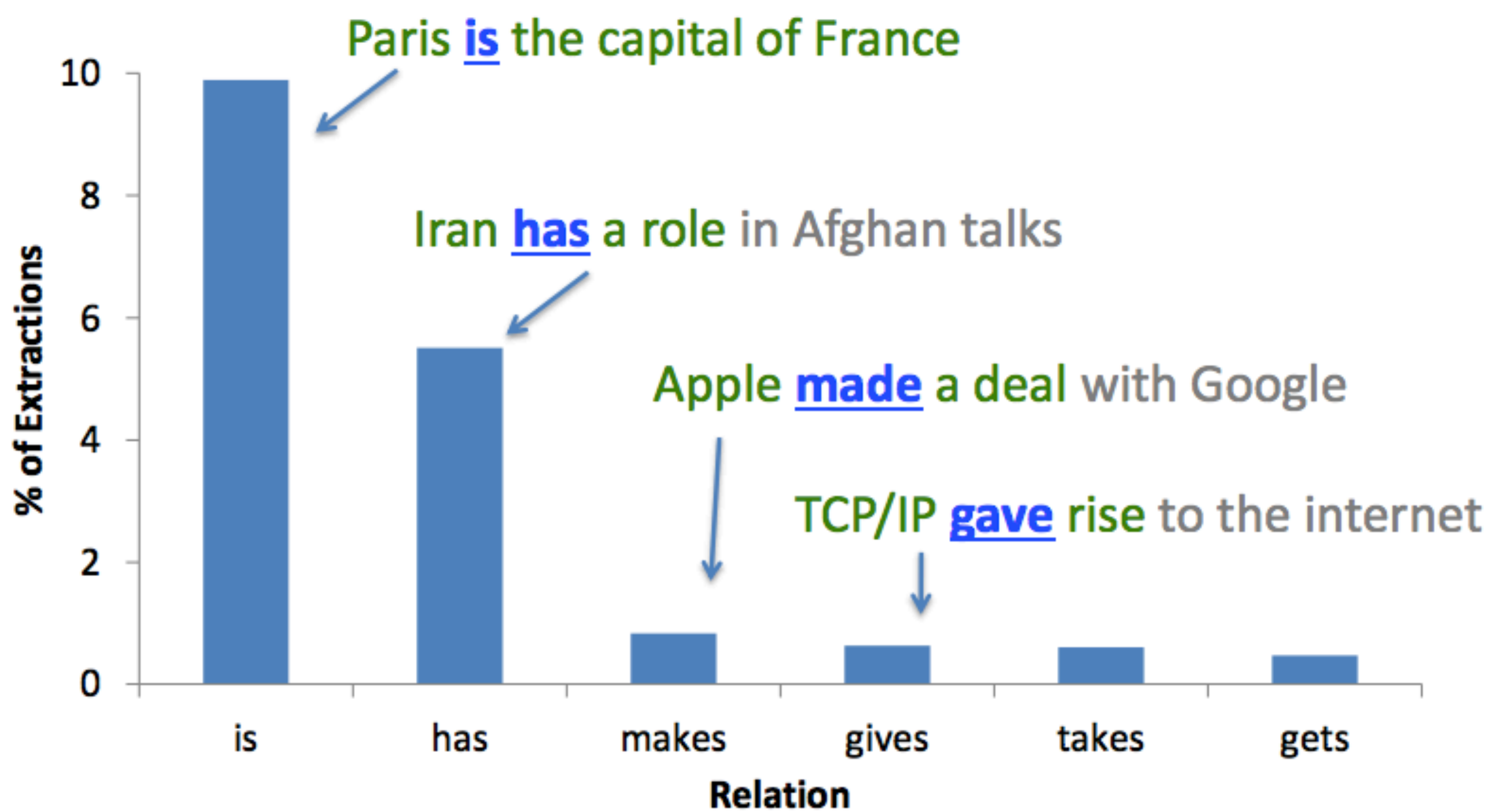
(**Arbor**, **for assumed**, **debt**).

Uninformative Extractions

Homer made a deal with the devil.

(**Homer**, **made**, **deal**)

Relation Frequency



A large proportion of relations appear in a handful of ways.
Let's focus on how relation phrases are specified!

ReVerb: Relation Extraction from Verbs

1. Identify longest relation phrases satisfying constraints

Hudson was born in Hampstead, which is a suburb of London.

arg1 arg2 arg1 arg2

2. Heuristically identify arguments for reach relation phrase

(Hudson, was born in, Hampstead)

(Hampstead, is a suburb of, London)

ReVerb Pattern Phrases

1) Use syntactic constraints to specify relation phrases.

Three simple patterns:

V	discovered	V = verb | particle | adv
V P	died from	P = prep | partcle | inf. marker
V W* P	played a role in	W = noun | adj | adv | det | pron

Find longest phrase matching one of the syntactic constraints.

2) Find nearest noun-phrases to the left and right of relation phrase.

- Not a relative pronoun or WHO-adverb or an existential there.

Lexical constraints for ReVerb

Problem: "overspecified" relation phrases

Obama is offering only modest greenhouse gas reduction targets at the conference.

Lexical constraints for ReVerb

Problem: "overspecified" relation phrases

Obama is offering only modest greenhouse gas reduction targets at the conference.

Solution: must have many distinct args in a large corpus

is offering only modest ...

Obama the conference ≈ 1

is the patron saint of

100s ≈

Anne mothers

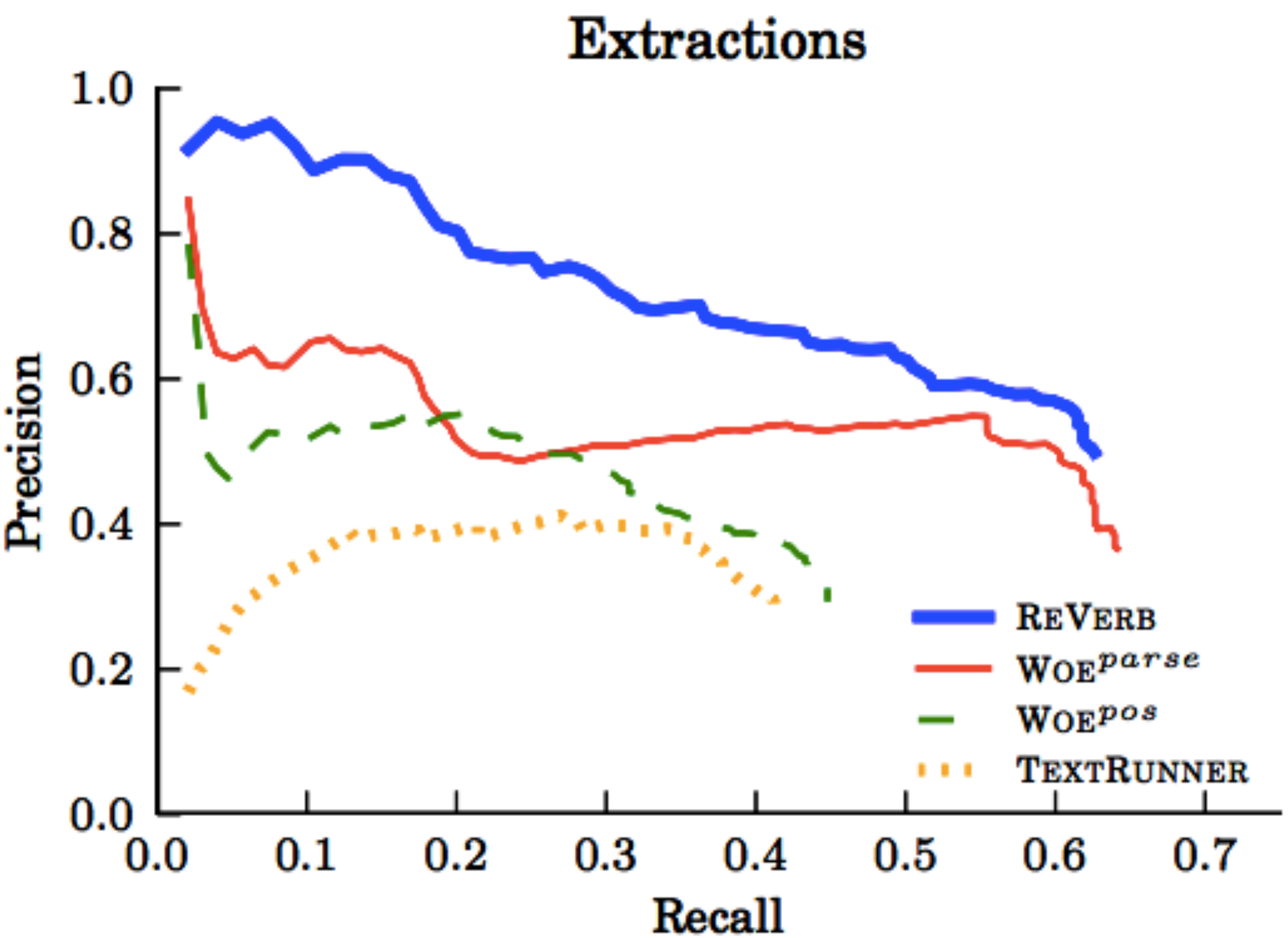
George England

Hubbins quality footwear

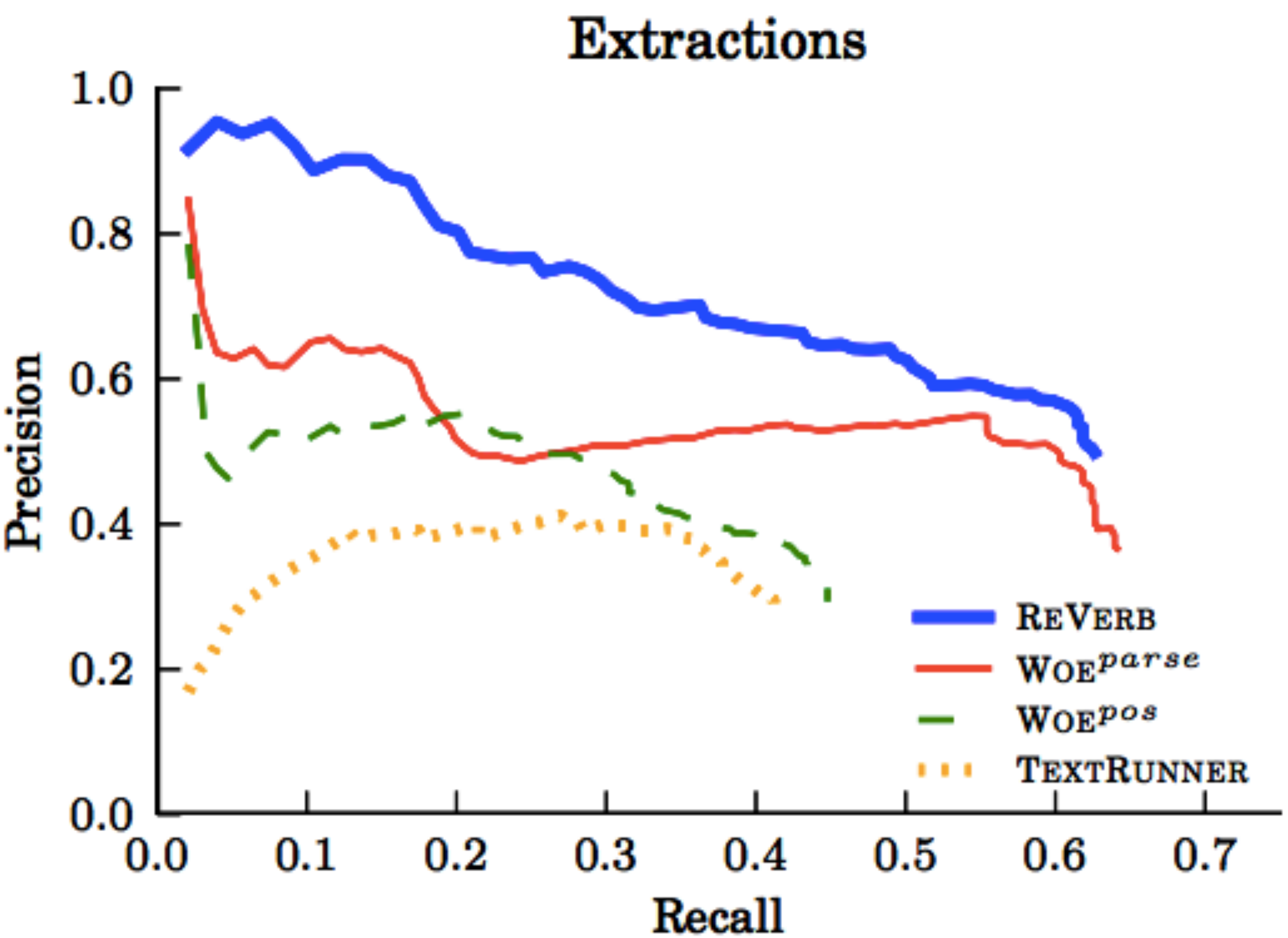
....

How good is ReVerb?

How good is ReVerb?

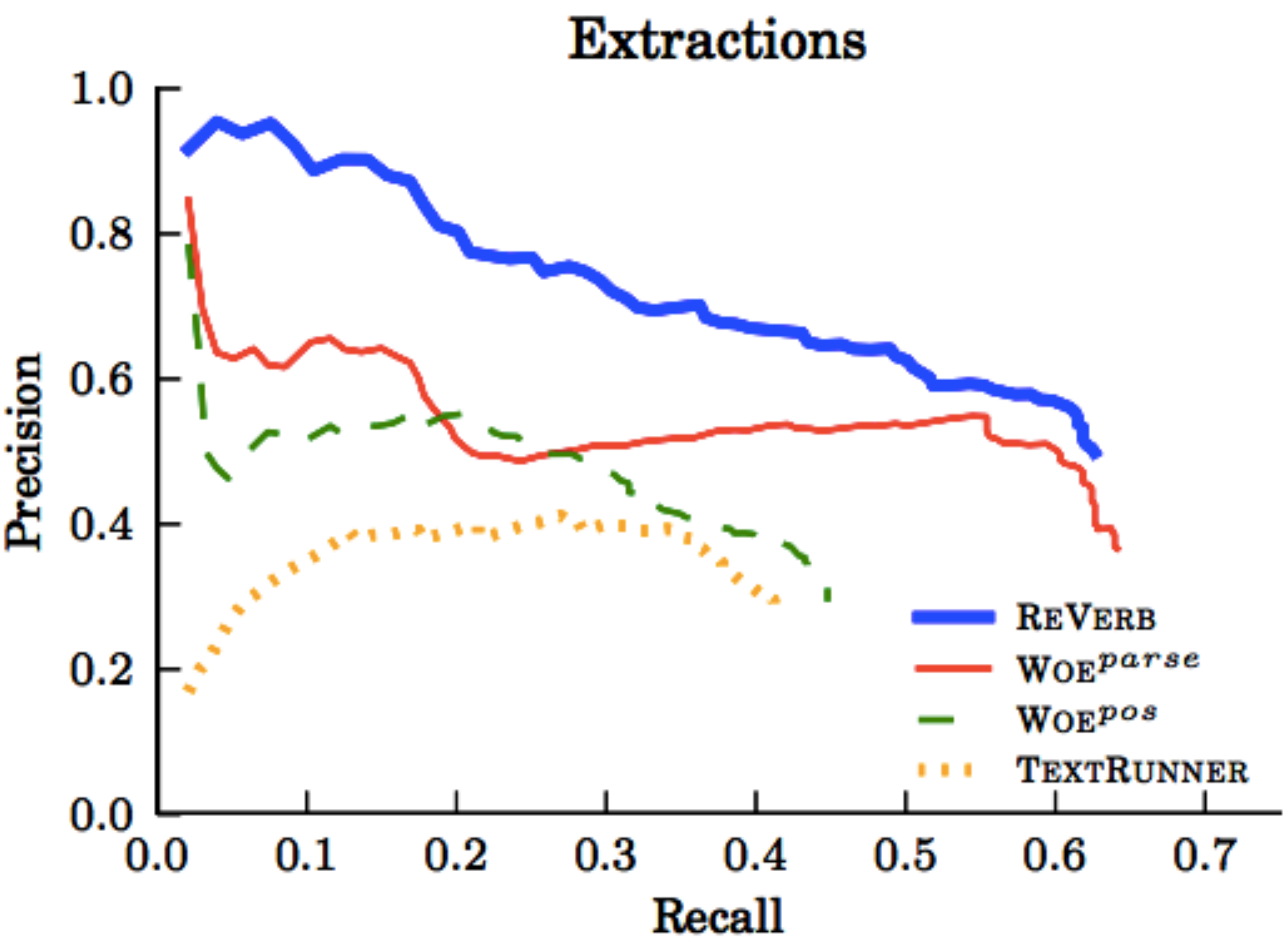


How good is ReVerb?



inhibits tumor growth in	**has a PhD in**	**joined forces with**
is a person who studies	**voted in favor of**	**won an Oscar for**
has a maximum speed of	**died from complications of**	**mastered the art of**
gained fame as	**granted political asylum to**	**is the patron saint of**

How good is ReVerb?



inhibits tumor growth in	has a PhD in	joined forces with
is a person who studies	voted in favor of	won an Oscar for
has a maximum speed of	died from complications of	mastered the art of
gained fame as	granted political asylum to	is the patron saint of

DARPA MR Domains	<50
NYU, Yago	<100
NELL	~500
DBpedia 3.2	940
PropBank	3,600
VerbNet	5,000
WikiPedia InfoBoxes, f > 10	~5,000
TextRunner	**100,000+**
ReVerb	**1,500,000+**

Key Issues

Key Issues

ReVerb - Incorrect Extractions	
65%	Correct relation phrase, incorrect arguments
16%	N-ary relation
8%	Non-contiguous relation phrase
2%	Imperative verb
2%	Overspecified relation phrase
7%	Other, including POS/chunking errors

Key Issues

ReVerb - Incorrect Extractions	
65%	Correct relation phrase, incorrect arguments
16%	N-ary relation
8%	Non-contiguous relation phrase
2%	Imperative verb
2%	Overspecified relation phrase
7%	Other, including POS/chunking errors

ReVerb - Missed Extractions	
52%	Could not identify correct arguments
23%	Relation filtered out by lexical constraint
17%	Identified a more specific relation
8%	POS/chunking error

Key Issues

ReVerb - Incorrect Extractions	
65%	Correct relation phrase, incorrect arguments
16%	N-ary relation
8%	Non-contiguous relation phrase
2%	Imperative verb
2%	Overspecified relation phrase
7%	Other, including POS/chunking errors

ReVerb - Missed Extractions	
52%	Could not identify correct arguments
23%	Relation filtered out by lexical constraint
17%	Identified a more specific relation
8%	POS/chunking error

Argument detection heuristic not adequate.
Lexical constraint too restrictive.

Arg Learning: What's Arg1?

Category	Pattern	Freq
Basic Noun Phrases Chicago was founded in 1833	NN, JJ NN, etc	65%
Prepositional Attachments The forest in Brazil is threatened by ranching.	NP PP NP	19%
List Google and Apple are headquartered in Silicon Valley.	NP, (NP,)* CC NP	15%
Relative Clause Chicago, which is located in Illinois, has three million residents.	NP (that|WP|WDT)? NP? VP NP	<1%

[Christensen et al, 2012]

Arg Learning: What's Arg2?

Category	Pattern	Freq
Basic Noun Phrases Calcium prevents osteoporosis	NN, JJ NN, etc	60%
Prepositional Attachments Barack Obama is one of the presidents of the United States	NP PP NP	18%
List A galaxy consists of stars and stellar remnants	NP, (NP,)* CC NP	15%
Independent Clause Scientists estimate that 80% of oil remains a threat.	(that|WP|WDT)? NP? VP NP	8%
Relative Clause The shooter killed a woman who was running from the scene.	NP (that|WP|WDT)? NP? VP NP	6%

[Christensen et al, 2012]

Arg Learner using CRFs

1) Build three classifiers

- Identify arg1 right bound
  ... TOK TOK TOK TOK TOK rel TOK TOK TOK ...
- Identify arg1 left bound
  ... TOK TOK TOK TOK TOK rel TOK TOK TOK ...
- Identify arg2 left bound
  ... TOK TOK TOK TOK TOK rel TOK TOK TOK ...
- Identify arg2 right bound
  ... TOK TOK TOK TOK TOK rel TOK TOK TOK ...

Classifier

Classifier

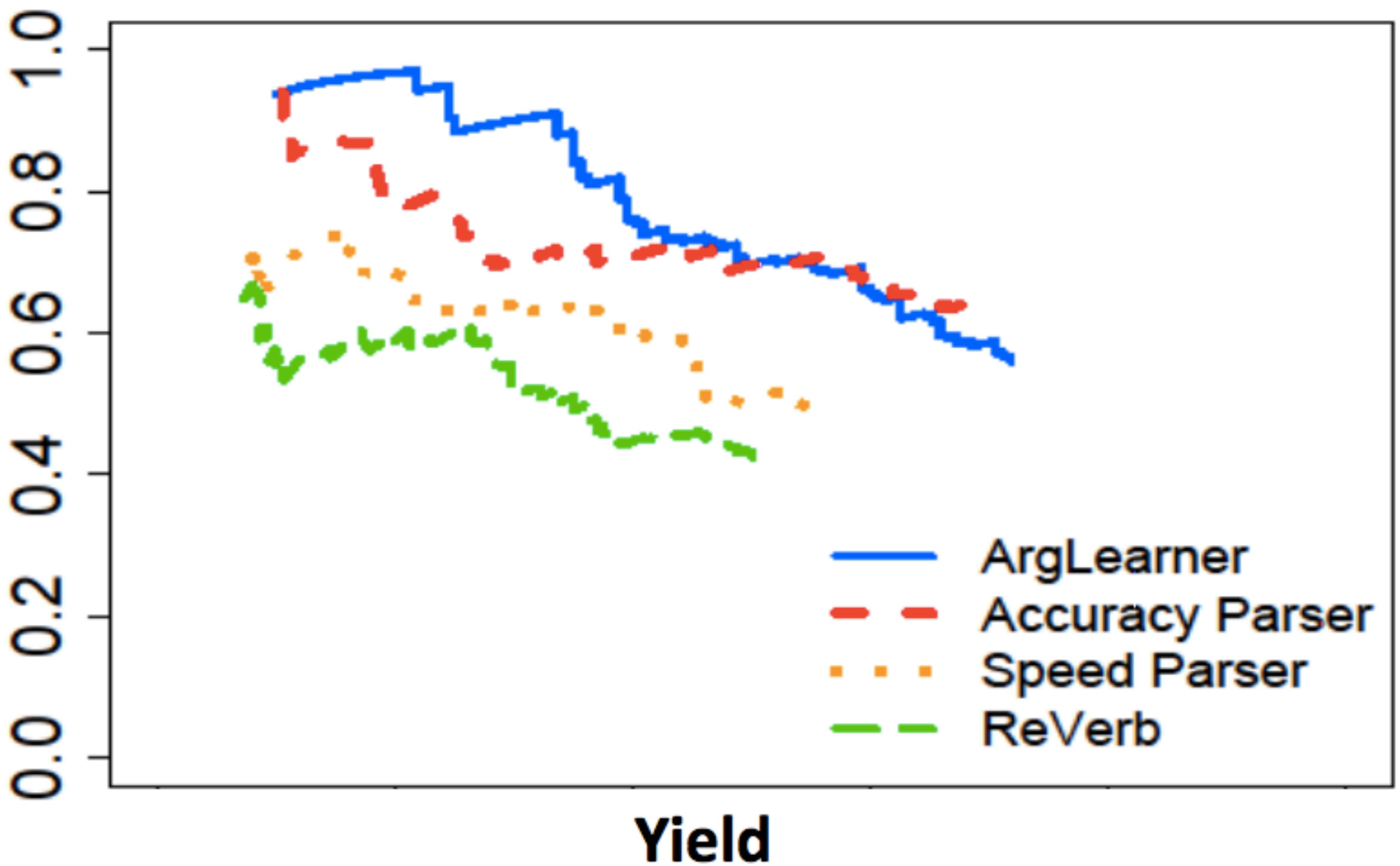
Classifier

2) Each with its own feature set based on the syntactic analysis

--

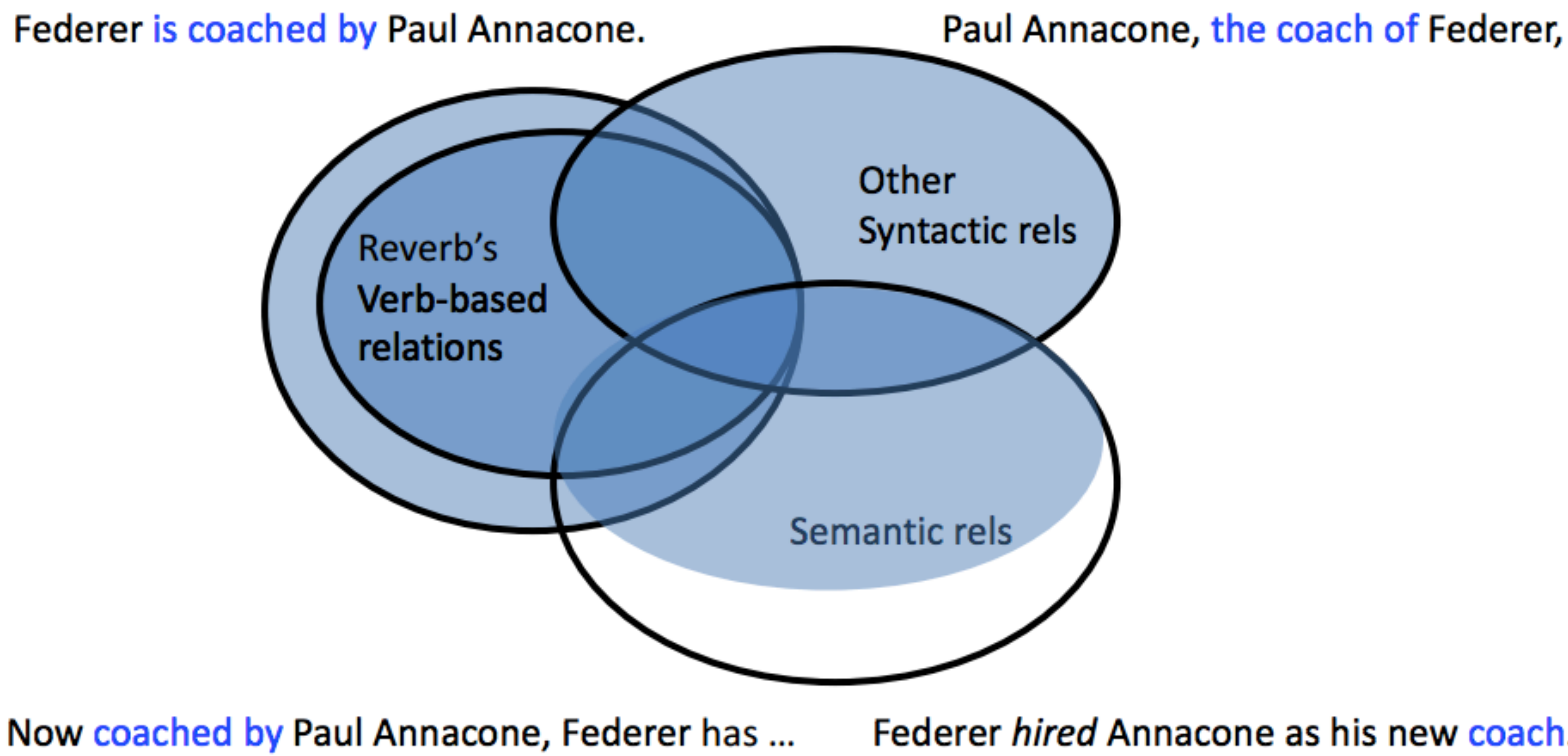
[Christensen et al, 2012]

Arg Learner

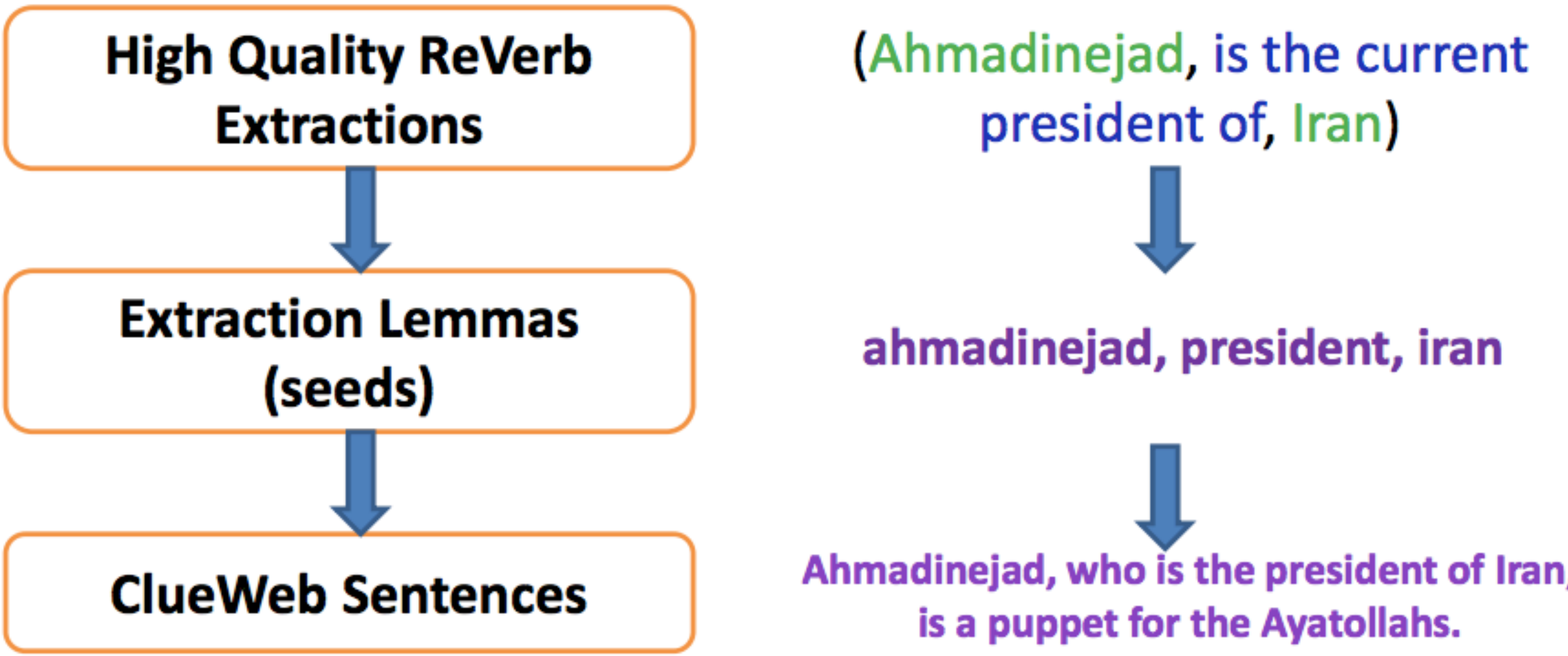


1.0
0.8
0.6
0.4
0.2
0.0
Yield
ArgLearner
Accuracy Parser
Speed Parser
ReVerb

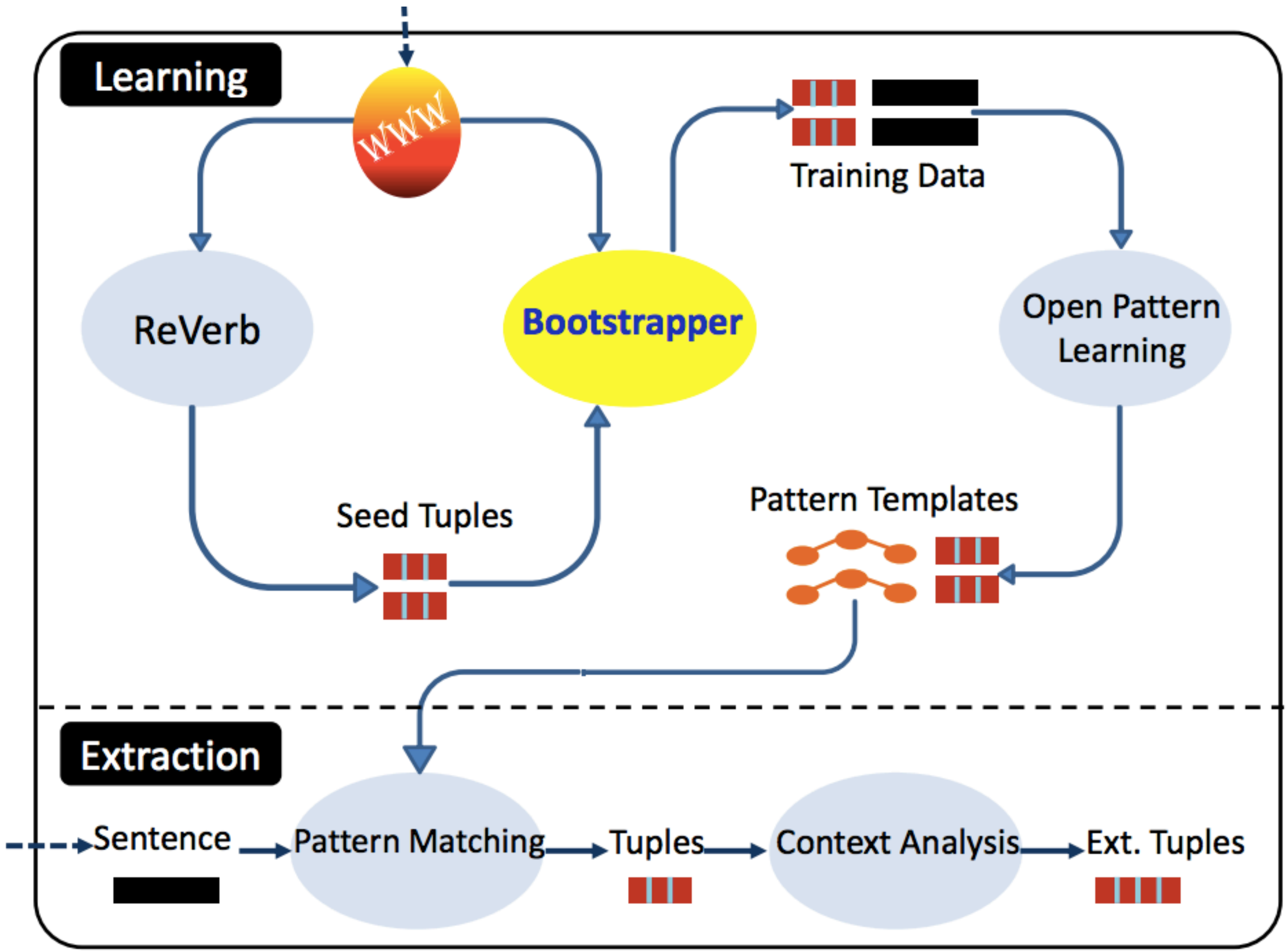
Ollie: Bootstrapping from ReVerb [Mausam et al., 2013]



Ollie: Bootstrapping from ReVerb



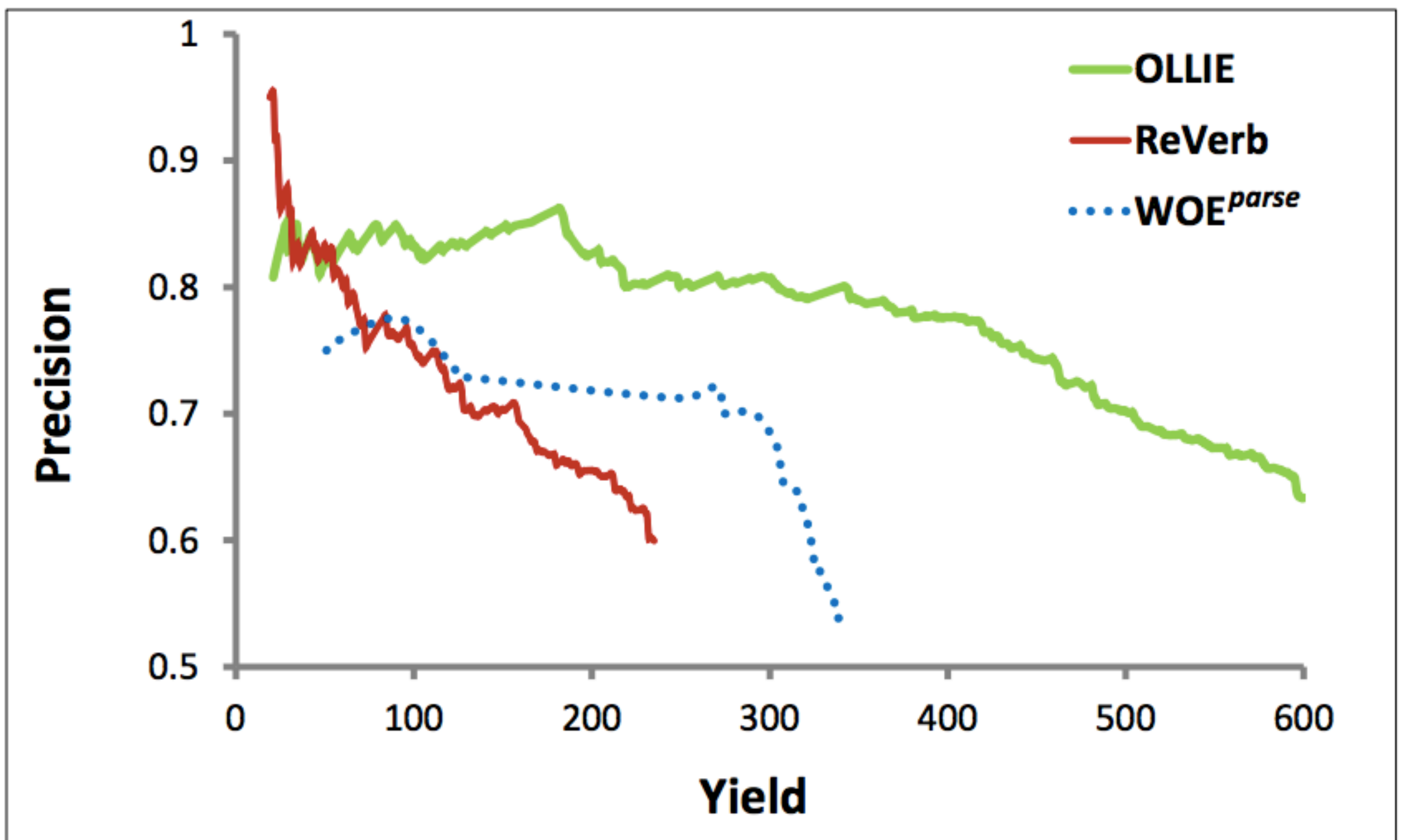
Ollie: Bootstrapping from ReVerb



Supervised Learning of Patterns

Features

– Frequency of pattern in training set
– Lexical/POS features
– Length/coverage features
– …



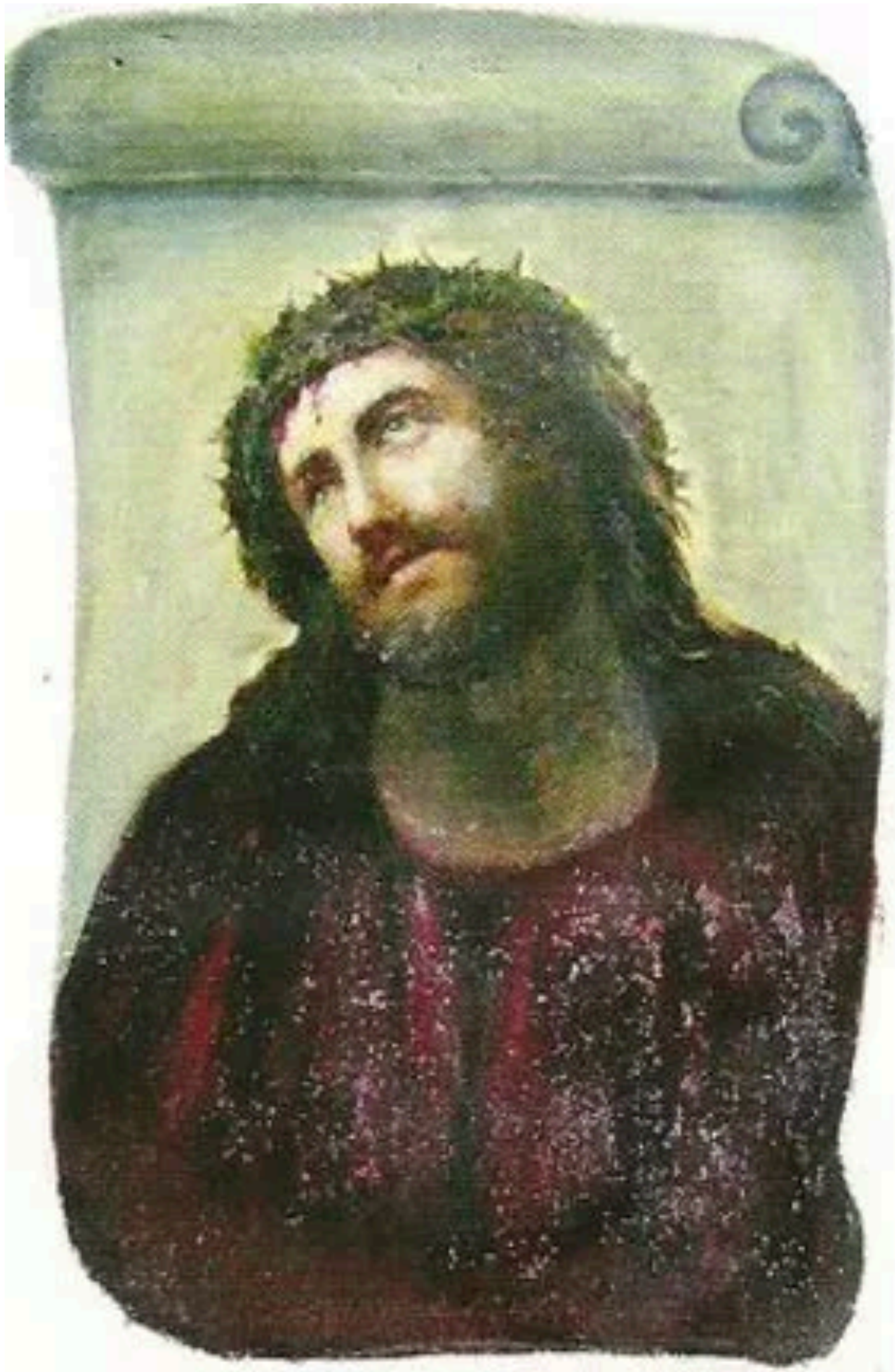
Issues with Open IE

Issues with Open IE

- **Semantics**?
  - Not tied to any ontology. Can't assign specific meaning.
  - Tie relation phrases back to an existing ontology [Soderland, 2012]
  - Learn inference rules over Open IE relations directly!

Issues with Open IE

- **Semantics**?
  - Not tied to any ontology. Can't assign specific meaning.
  - Tie relation phrases back to an existing ontology [Soderland, 2012]
  - Learn inference rules over Open IE relations directly!
- **Redundancy**
  - Many distinct relation phrases convey the same conceptual "relation"
  - Solution: Cluster relations



Entity Linking

Person Name Disambiguation


Computational Linguist

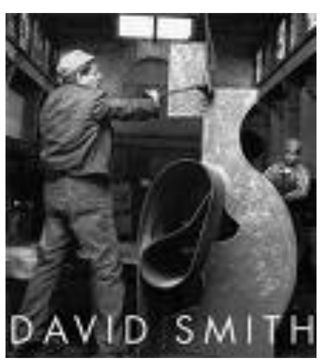

Musician


Physicist


Psychologist


Photographer


CEO



Sculptor


Biologist

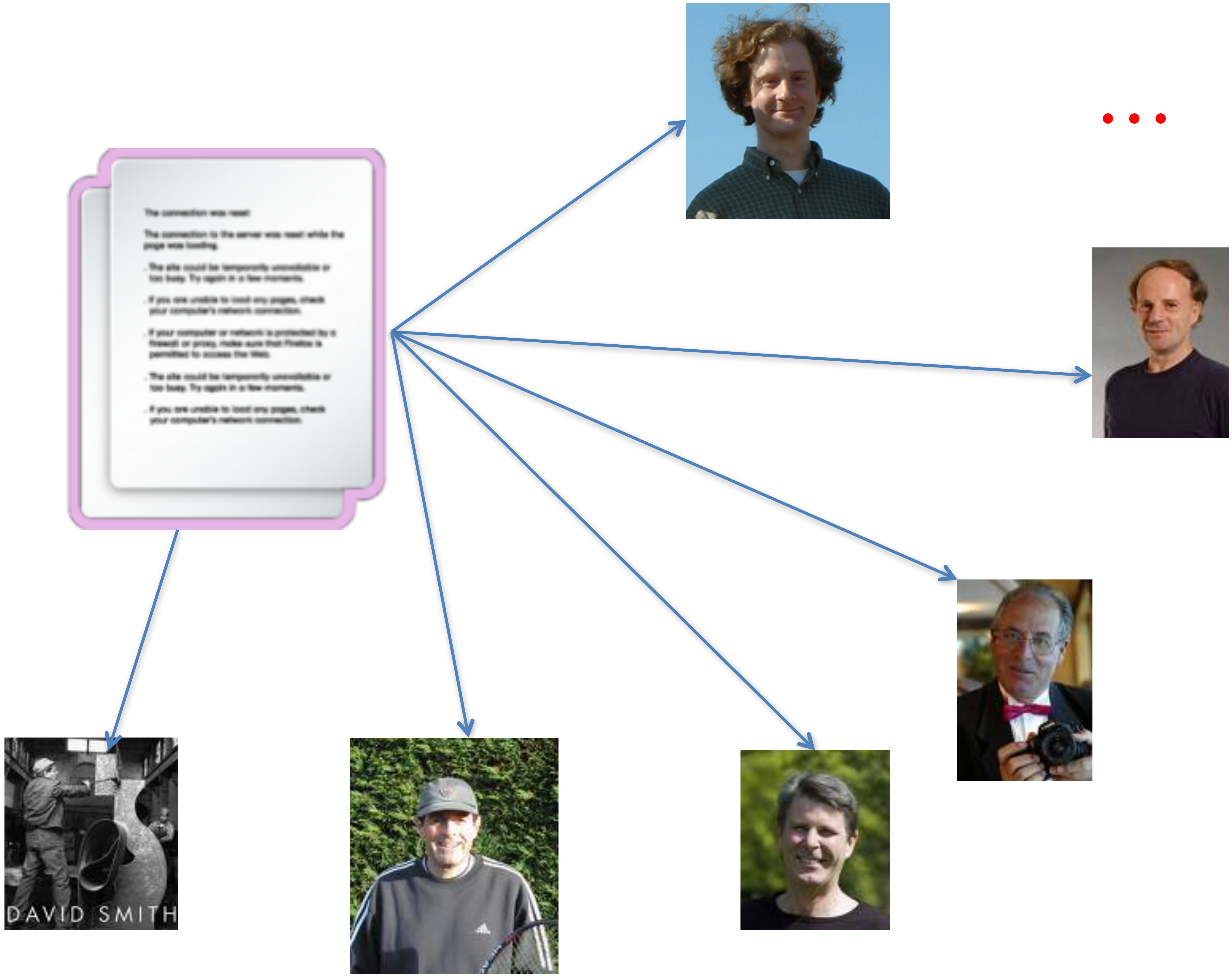

Tennis Player

. . .


Pastor

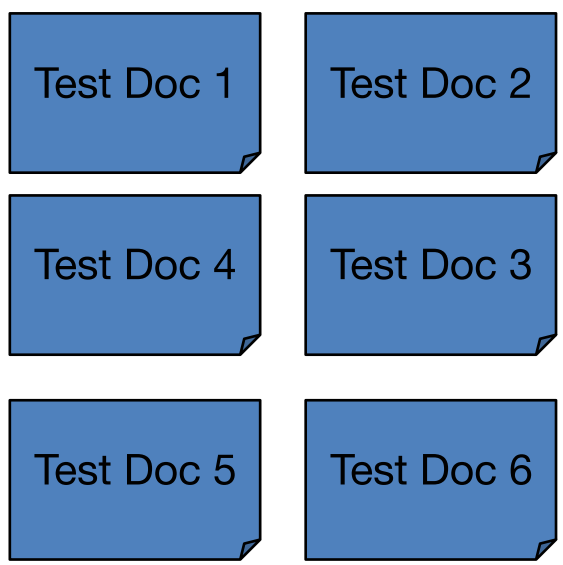

Theologist

Rao, Garera & Yarowsky, 2007

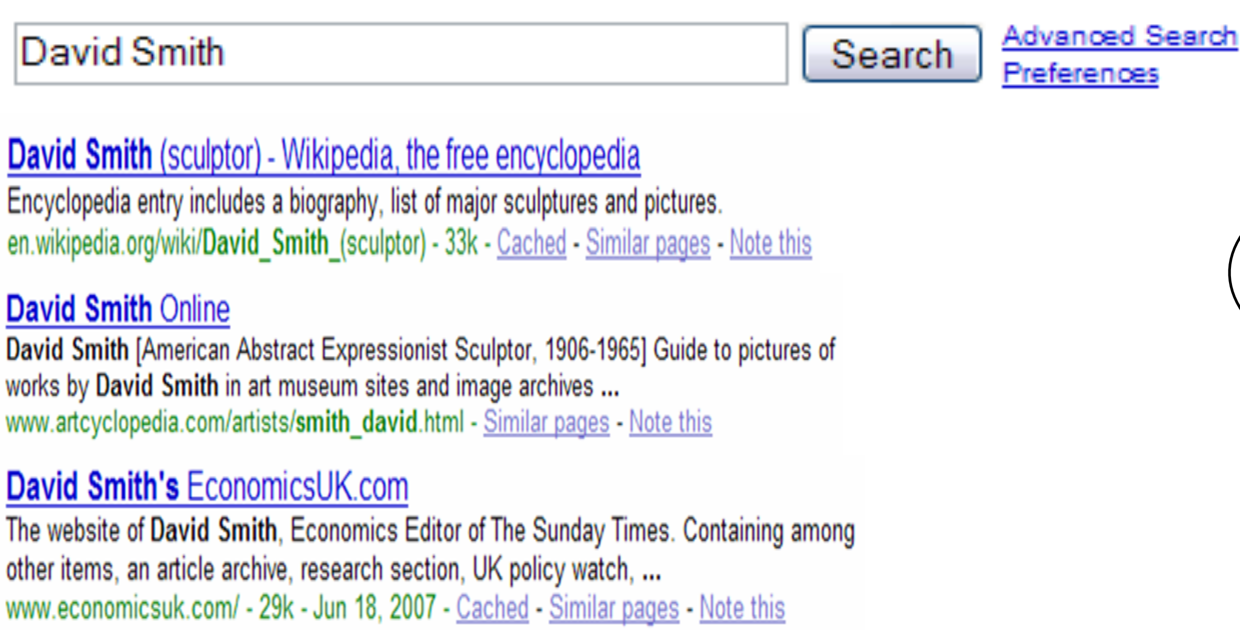

DAVID SMITH

Clustering using web snippets

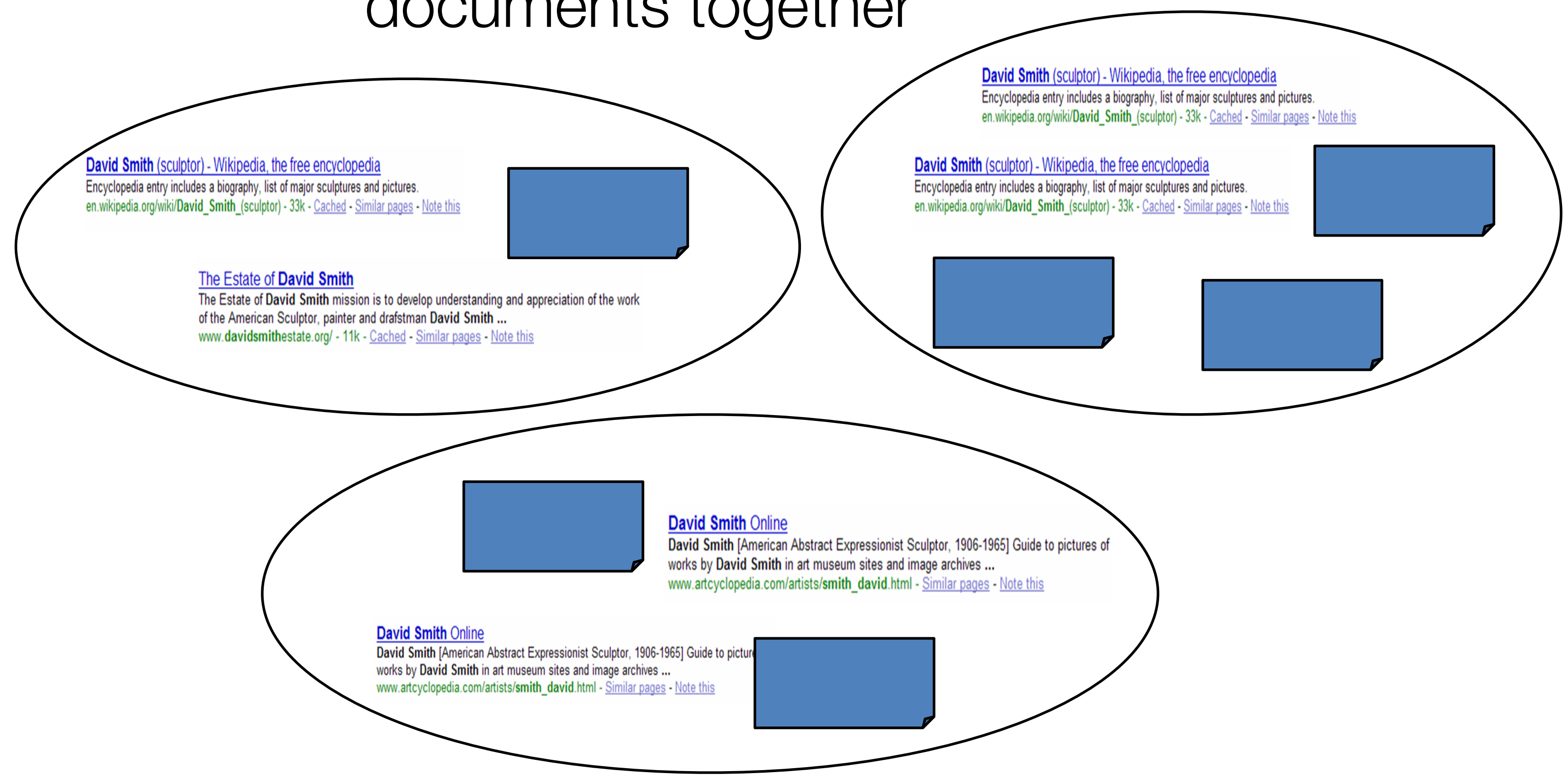
Goal: To cluster 100 given test documents for name “David Smith”



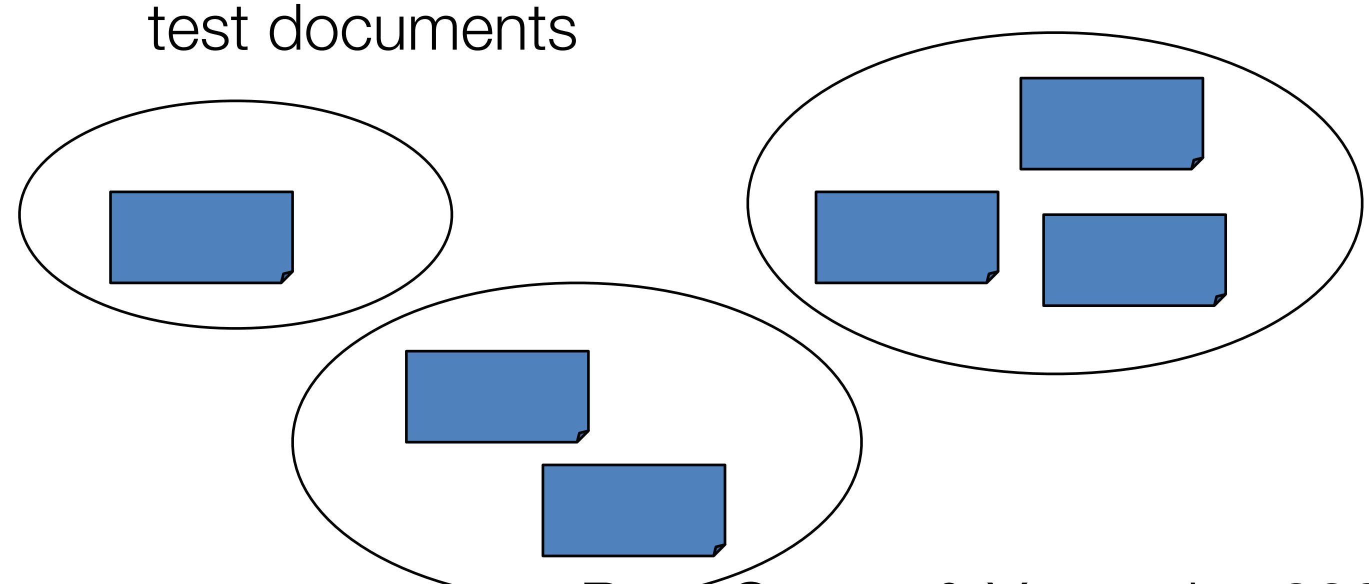
Step 1: Extract top 1000 snippets from Google



Step 2: Cluster all the 1100 documents together



Step 3: Extract the clustering of the test documents



Rao, Garera & Yarowsky, 2007

Web Snippets for Disambiguation

Snippet

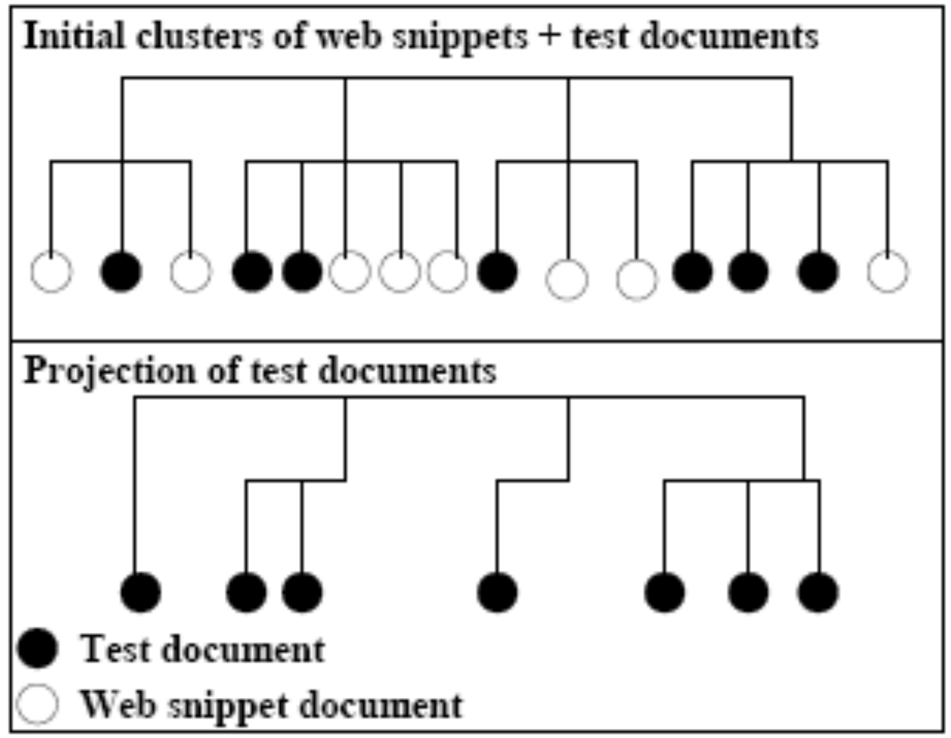
Dekang Lin's Home Page
Dekang Lin, Professor. Department of Computing Science · University of Alberta. Edmonton, Alberta, Canada, T6G 2H1. Phone: 780 492-9920. Fax: 780 492-1071 **...**
www.cs.ualberta.ca/~**lin**dek/ - 12k - Cached - Similar pages - Note this

Downloads
Minipar, Minipar is a principle-based broad coverage parser. The version that is downloadable from here contains. HMM, A HMM Package in C++. **...**
www.cs.ualberta.ca/~**lin**dek/downloads.htm - 6k - Cached - Similar pages - Note this
[More results from www.cs.ualberta.ca]

- Snippets contain high quality, low noise features
- Easy to extract
- Derived from sources other than the document (e.g., link text)

Rao, Garera & Yarowsky, 2007

Term bridging via Snippets



Document 1

Contains term
"780 492-9920"

Document 2

Contains term
"T6G2H1"

Dekang Lin's Home Page
Dekang Lin, Professor. Department of Computing Science · University of Alberta. Edmonton, Alberta, Canada, T6G 2H1. Phone: 780 492-9920. Fax: 780 492-1071 **...**
www.cs.ualberta.ca/~**lindek**/ - 12k - Cached - Similar pages - Note this

Snippet contains both the terms "780 492-9920" "T6G2H1" and that can serve as a bridge for clustering Document 1 and Document 2 together

Rao, Garera & Yarowsky, 2007

Entity Linking

John Williams

Richard Kaufman goes a long way back with **John Williams**. Trained as a classical violinist, Californian Kaufman started doing session work in the Hollywood studios in the 1970s. One of his movies was Jaws, with **Williams** conducting his score in recording sessions in 1975...

John Williams	author	1922-1994
J. Lloyd Williams	botanist	1854-1945
John Williams	politician	1955-
John J. Williams	US Senator	1904-1988
John Williams	Archbishop	1582-1650
John Williams	composer	1932-
Jonathan Williams	poet	1929-

Michael Phelps

Debbie Phelps, the mother of swimming star **Michael Phelps**, who won a record eight gold medals in Beijing, is the author of a new memoir, ...

Michael Phelps is the scientist most often identified as the inventor of PET, a technique that permits the imaging of biological processes in the organ systems of living individuals. **Phelps** has ...

Michael Phelps	swimmer	1985-
Michael Phelps	biophysicist	1939-

Identify matching entry, or determine that entity is missing from KB

Challenges in Entity Linking

Challenges in Entity Linking

- Name Variation
  - Abbreviations: BSO vs. Boston Symphony Orchestra
  - Shortened forms: Osama Bin Laden vs. Bin Laden
  - Alternate spellings:
    Osama vs. Ussamah vs. Oussama

Challenges in Entity Linking

- Name Variation
  - Abbreviations: BSO vs. Boston Symphony Orchestra
  - Shortened forms: Osama Bin Laden vs. Bin Laden
  - Alternate spellings:
    Osama vs. Ussamah vs. Oussama
- Entity Ambiguity: Polysemous mentions
  - E.g., Springfield, Washington

Challenges in Entity Linking

- Name Variation
  - Abbreviations: BSO vs. Boston Symphony Orchestra
  - Shortened forms: Osama Bin Laden vs. Bin Laden
  - Alternate spellings:
    Osama vs. Ussamah vs. Oussama
- Entity Ambiguity: Polysemous mentions
  - E.g., Springfield, Washington
- Absence: Open domain linking
  - Not all observed mentions have a corresponding entry in KB (NIL mentions)
  - Ability to predict NIL mentions determines KBP accuracy
  - Largely overlooked in current literature

Entity Linking: Features

Entity Linking: Features

- Name-matching
  - acronyms, aliases, string-similarity, probabilistic FST

Entity Linking: Features

- Name-matching
  - acronyms, aliases, string-similarity, probabilistic FST
- Document Features
  - TF/IDF comparisons, occurrence of names or KB facts in the query text, Wikitology

Entity Linking: Features

- Name-matching
  - acronyms, aliases, string-similarity, probabilistic FST
- Document Features
  - TF/IDF comparisons, occurrence of names or KB facts in the query text, Wikitology
- KB Node
  - Type (e.g., is this a person), Features of Wikipedia page, Google rank of corresponding Wikipedia page

Entity Linking: Features

- Name-matching
  - acronyms, aliases, string-similarity, probabilistic FST
- Document Features
  - TF/IDF comparisons, occurrence of names or KB facts in the query text, Wikitology
- KB Node
  - Type (e.g., is this a person), Features of Wikipedia page, Google rank of corresponding Wikipedia page
- Absence (NIL Indications)
  - Does any candidate look like a good string match?

Entity Linking: Features

- Name-matching
  - acronyms, aliases, string-similarity, probabilistic FST
- Document Features
  - TF/IDF comparisons, occurrence of names or KB facts in the query text, Wikitology
- KB Node
  - Type (e.g., is this a person), Features of Wikipedia page, Google rank of corresponding Wikipedia page
- Absence (NIL Indications)
  - Does any candidate look like a good string match?
- Combinations
  - Low-string-match AND Acronym AND Type-is-ORG

Entity Linking: Name Matching

Entity Linking: Name Matching

- Acronyms

Entity Linking: Name Matching

- Acronyms
- Alias Lists
  - Wikipedia redirects, stock symbols, misc. aliases

Entity Linking: Name Matching

- Acronyms
- Alias Lists
  - Wikipedia redirects, stock symbols, misc. aliases
- Exact Match
  - With and without normalized punctuation, case, accents, appositive removal

Entity Linking: Name Matching

- Acronyms
- Alias Lists
  - Wikipedia redirects, stock symbols, misc. aliases
- Exact Match
  - With and without normalized punctuation, case, accents, appositive removal
- Fuzzier Matching
  - Dice score (character uni/bi/tri-grams), Hamming, Recursive LCSubstring, Subsequences
  - Word removal (e.g., Inc., US) and abbrev. expansion

Entity Linking: Name Matching

- Acronyms
- Alias Lists
  - Wikipedia redirects, stock symbols, misc. aliases
- Exact Match
  - With and without normalized punctuation, case, accents, appositive removal
- Fuzzier Matching
  - Dice score (character uni/bi/tri-grams), Hamming, Recursive LCSubstring, Subsequences
  - Word removal (e.g., Inc., US) and abbrev. expansion
- Weighted FST for Name Equivalence
  - Trained models score name-1 as a re-writing of name-2

Entity Linking: Document Features

Entity Linking: Document Features

- BoW Comparisons
  - TF/IDF & Dice scores for news article and KB text
  - Examined entire articles and passages around query mentions

Entity Linking: Document Features

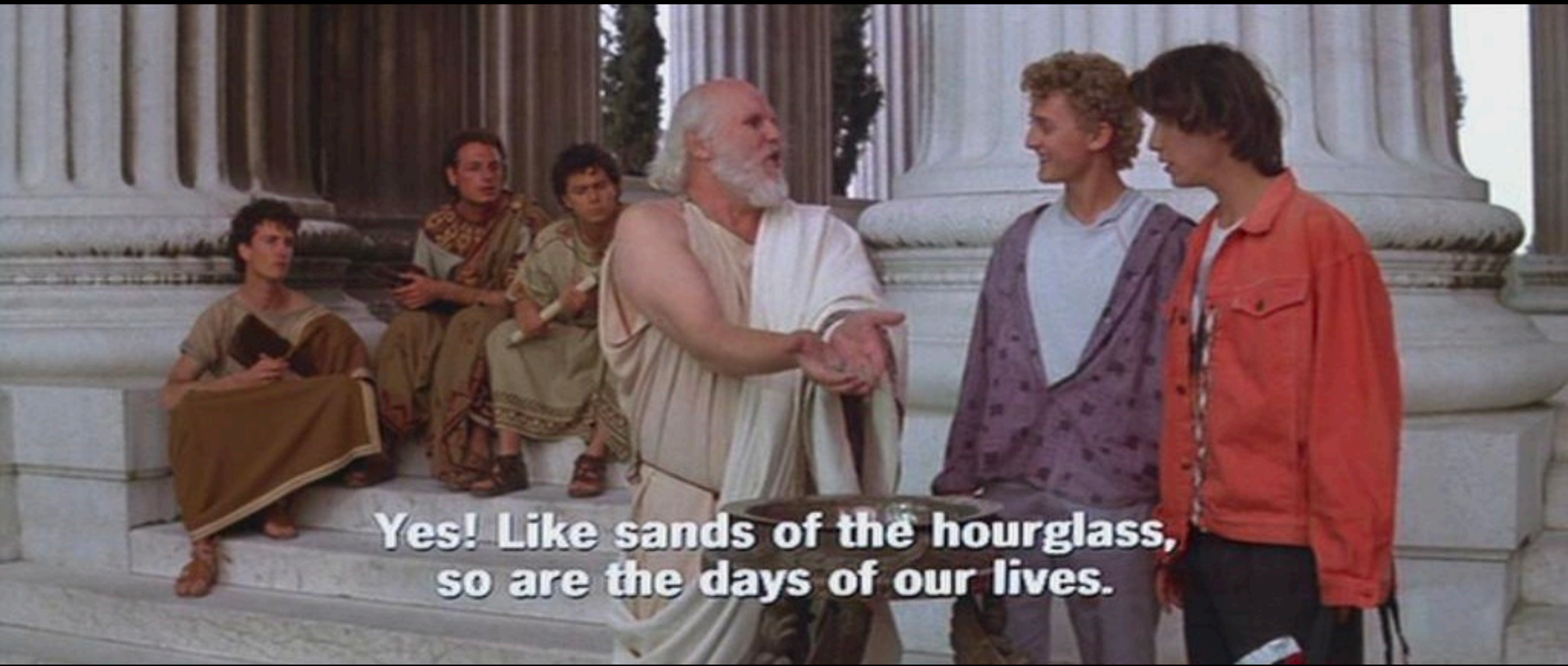
- BoW Comparisons
  - TF/IDF & Dice scores for news article and KB text
  - Examined entire articles and passages around query mentions
- Named-Entities
  - Ran BBN's SERIF analyzer on articles
  - Checked for coverage of (1) query co-references and (2) all names/nominals in KB text
  - Noted type, subtype of query entity (e.g., ORG/Media)

Entity Linking: Document Features

- BoW Comparisons
  - TF/IDF & Dice scores for news article and KB text
  - Examined entire articles and passages around query mentions
- Named-Entities
  - Ran BBN's SERIF analyzer on articles
  - Checked for coverage of (1) query co-references and (2) all names/nominals in KB text
  - Noted type, subtype of query entity (e.g., ORG/Media)
- KB Facts
  - Looked to see if candidate node's attributes are present in article text (e.g., spouse, employer, nationality)

Entity Linking: Document Features

- BoW Comparisons
  - TF/IDF & Dice scores for news article and KB text
  - Examined entire articles and passages around query mentions
- Named-Entities
  - Ran BBN's SERIF analyzer on articles
  - Checked for coverage of (1) query co-references and (2) all names/nominals in KB text
  - Noted type, subtype of query entity (e.g., ORG/Media)
- KB Facts
  - Looked to see if candidate node's attributes are present in article text (e.g., spouse, employer, nationality)
- Wikitology
  - UMBC system predicts relevant Wikipedia pages (or KB nodes) for text


Yes! Like sands of the hourglass,
so are the days of our lives.

Textual Entailment

The Holy grail of NLP….

- Understanding Natural Language Text

The Holy grail of NLP….

- Understanding Natural Language Text

  British left waffles on nukes

The Holy grail of NLP….

- Understanding Natural Language Text

    British left waffles on nukes

  - Traditional approach: map it to a canonical form
  - Can then (in theory) integrate multiple statements from diverse sources to derive “new” facts

The Holy grail of NLP….

- Understanding Natural Language Text

    British left waffles on nukes

  - Traditional approach: map it to a canonical form
  - Can then (in theory) integrate multiple statements from diverse sources to derive “new” facts
  - Question #1: How to represent its meaning?

The Holy grail of NLP….

- Understanding Natural Language Text

  British left waffles on nukes

  - Traditional approach: map it to a canonical form
  - Can then (in theory) integrate multiple statements from diverse sources to derive “new” facts
  - Question #1: How to represent its meaning?

$$\exists_x \exists_y \exists_z British(x) \wedge Waffles(y) \wedge Nukes(z) \wedge leave\_on(x,y,z)$$

  - Question #0.5: What *is* its meaning?

The Holy grail of NLP….

- Understanding Natural Language Text

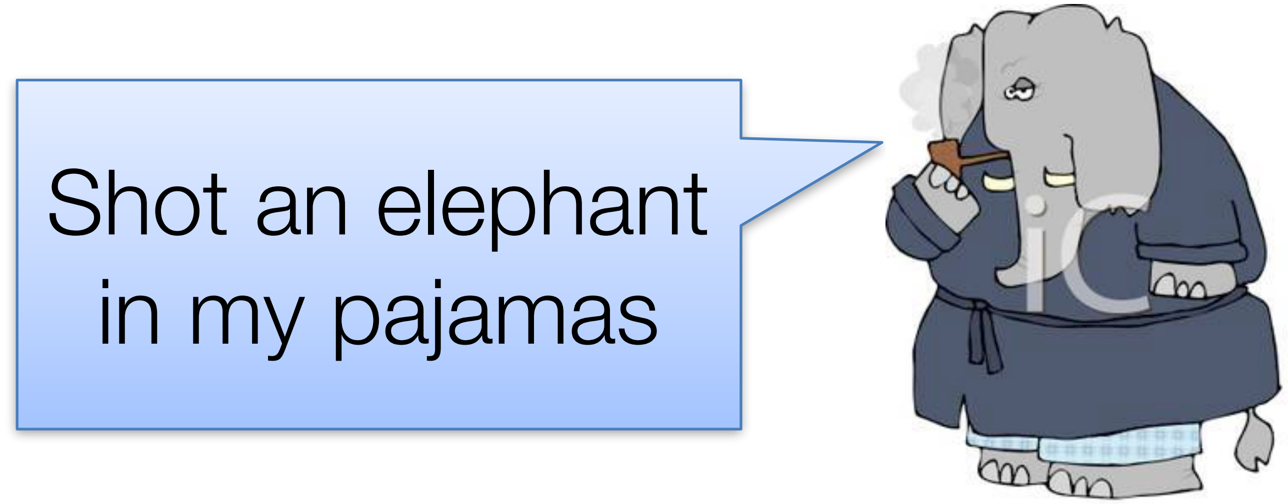
  British left waffles on nukes

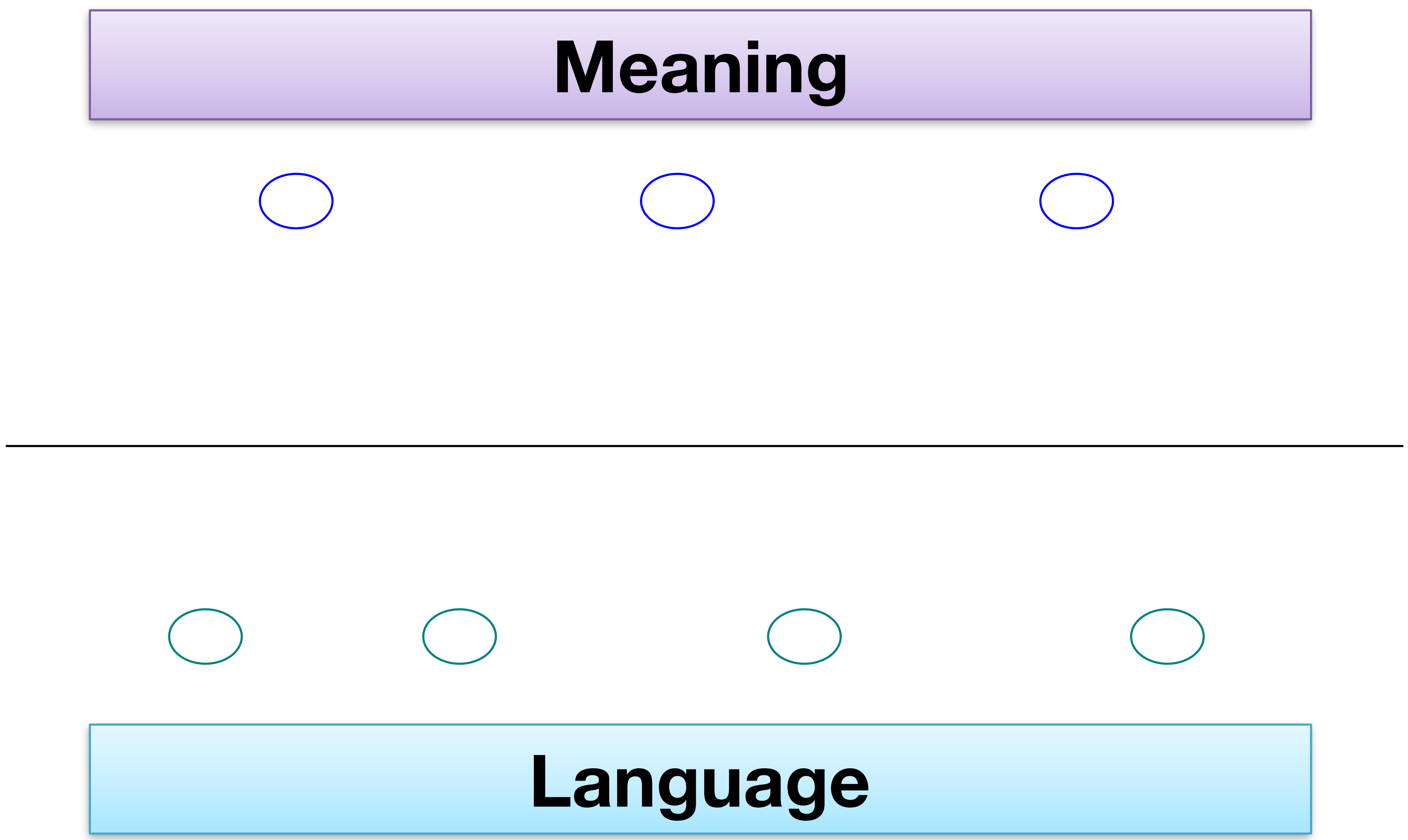
  - Traditional approach: map it to a canonical form
  - Can then (in theory) integrate multiple statements from diverse sources to derive "new" facts
  - Question #1: How to represent its meaning?

$$\exists_x \exists_y \exists_z British(x) \wedge Waffles(y) \wedge Nukes(z) \wedge leave\_on(x, y, z)$$

  - Question #0.5: What *is* its meaning?
  - Question #0.1: What does *understand* mean?

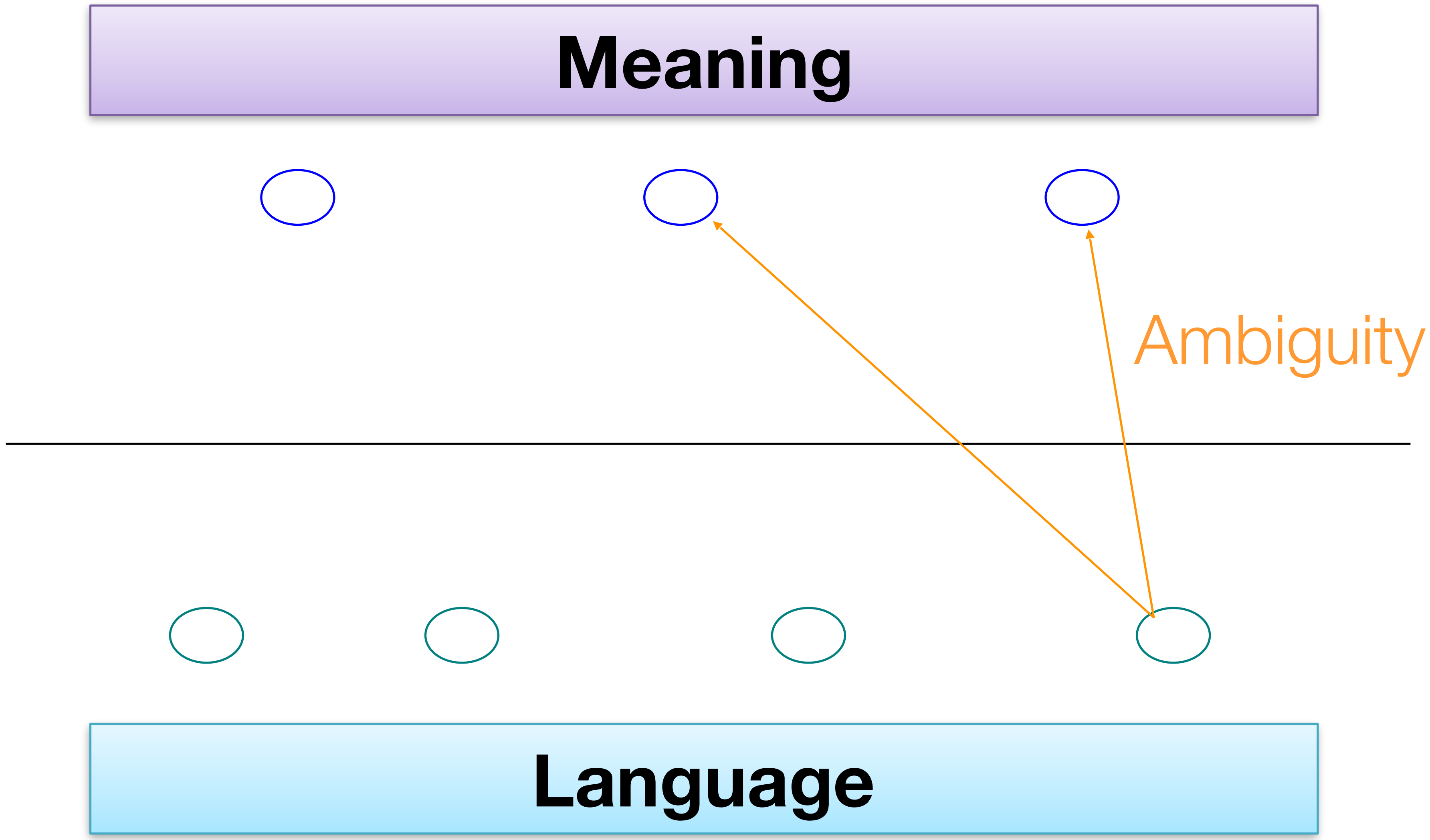
Natural Language and Meaning





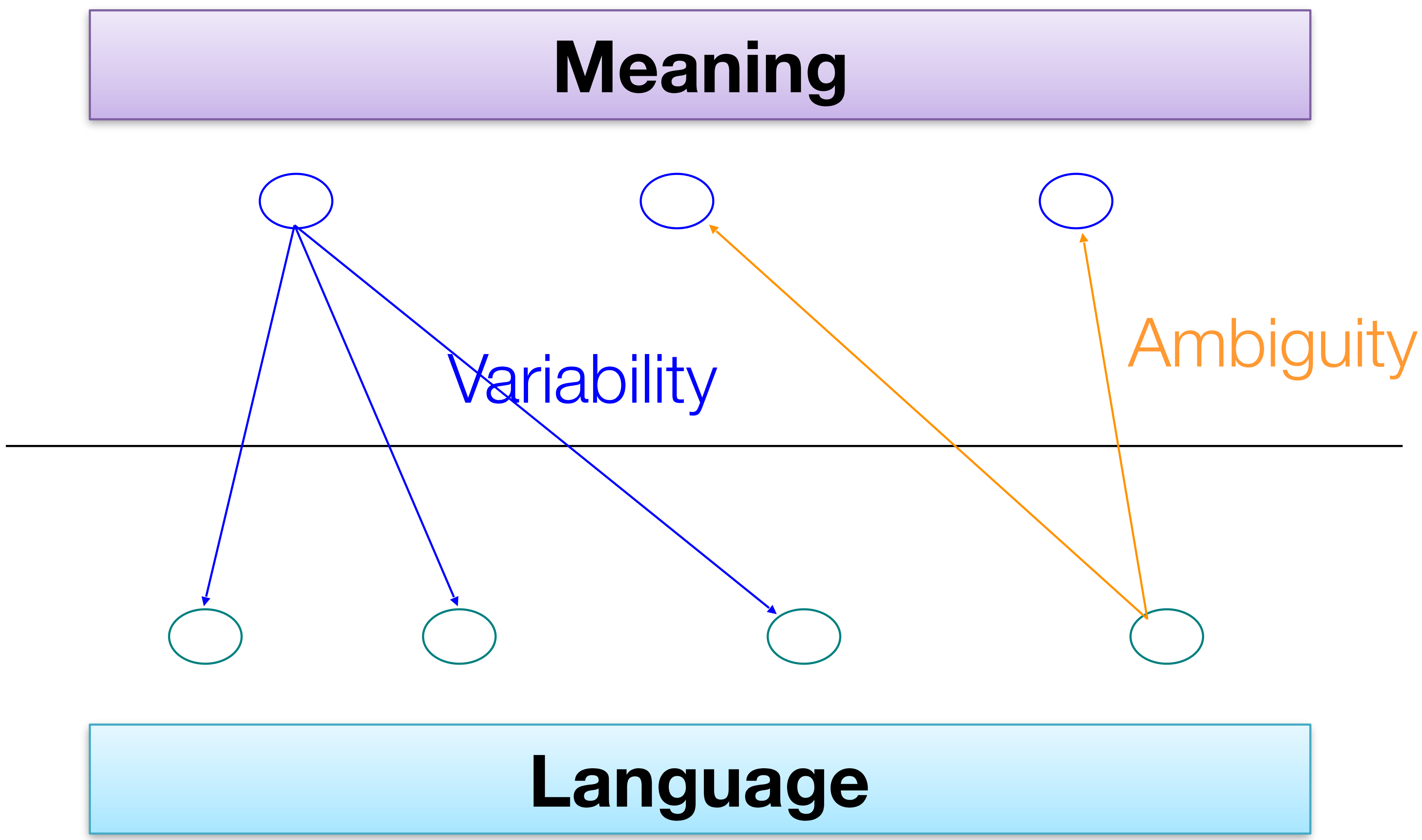
Natural Language and Meaning

Shot an elephant in my pajamas



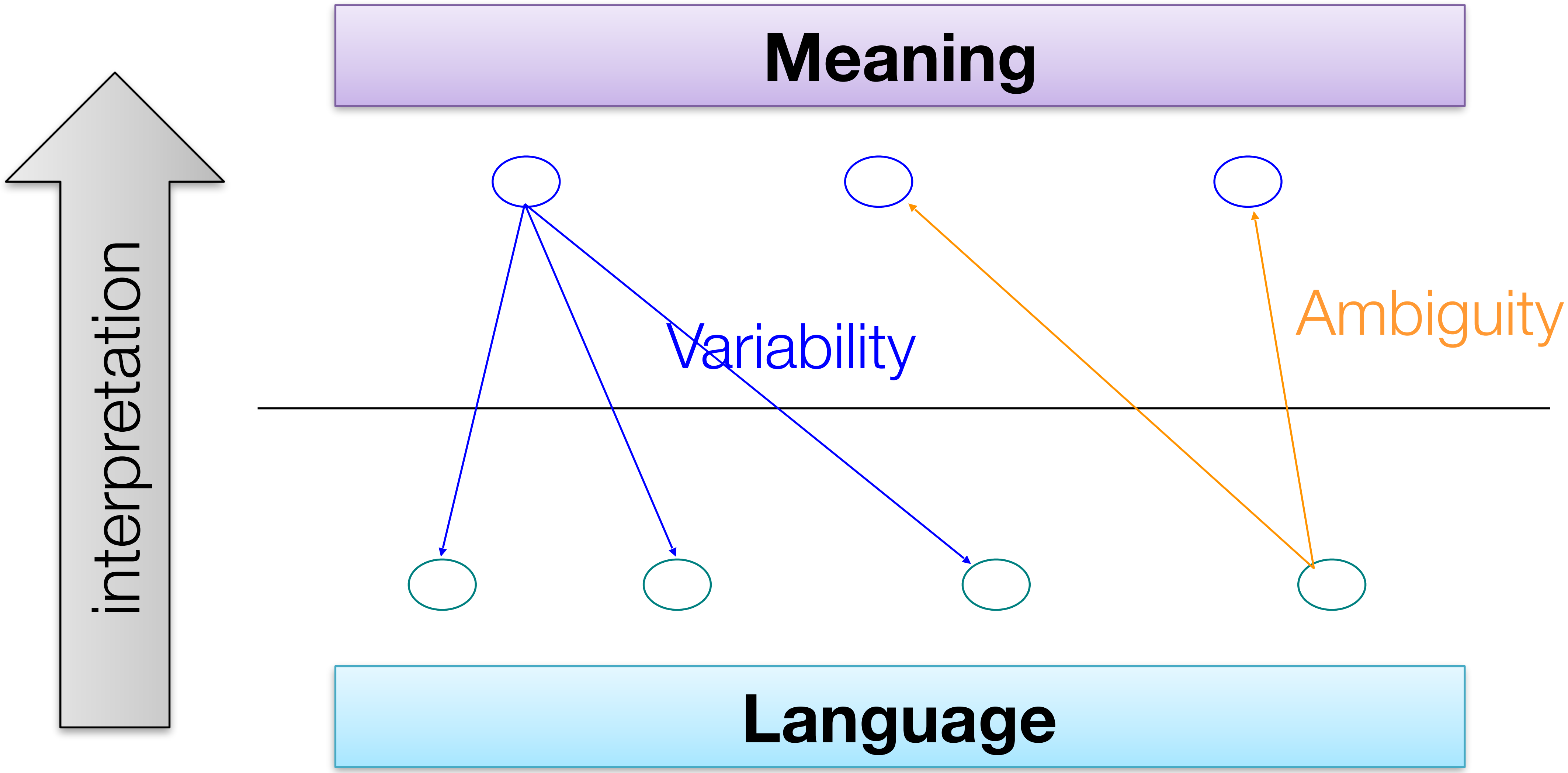
Natural Language and Meaning

Shot an elephant in my pajamas

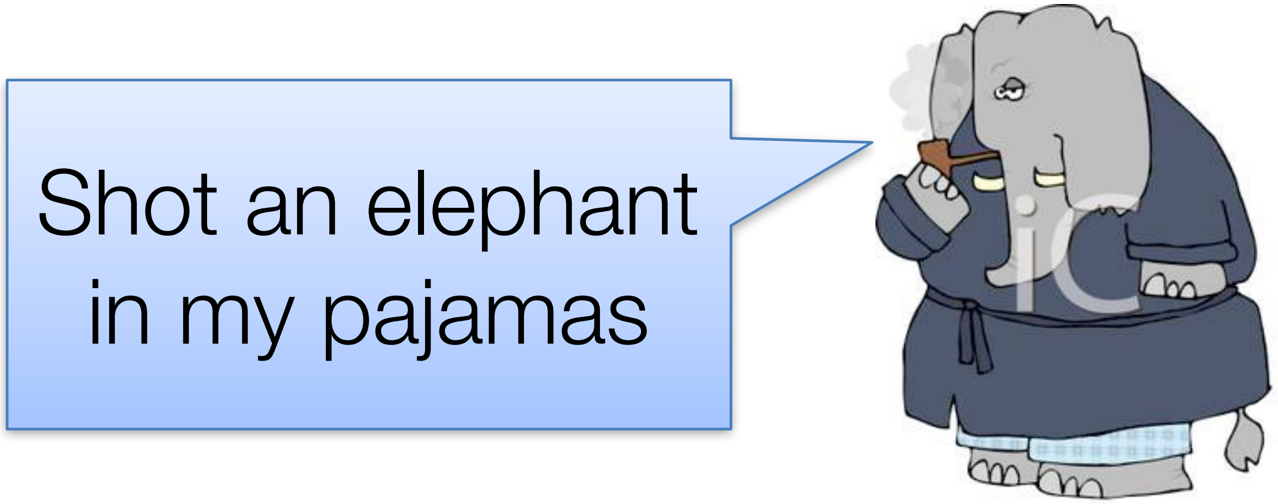


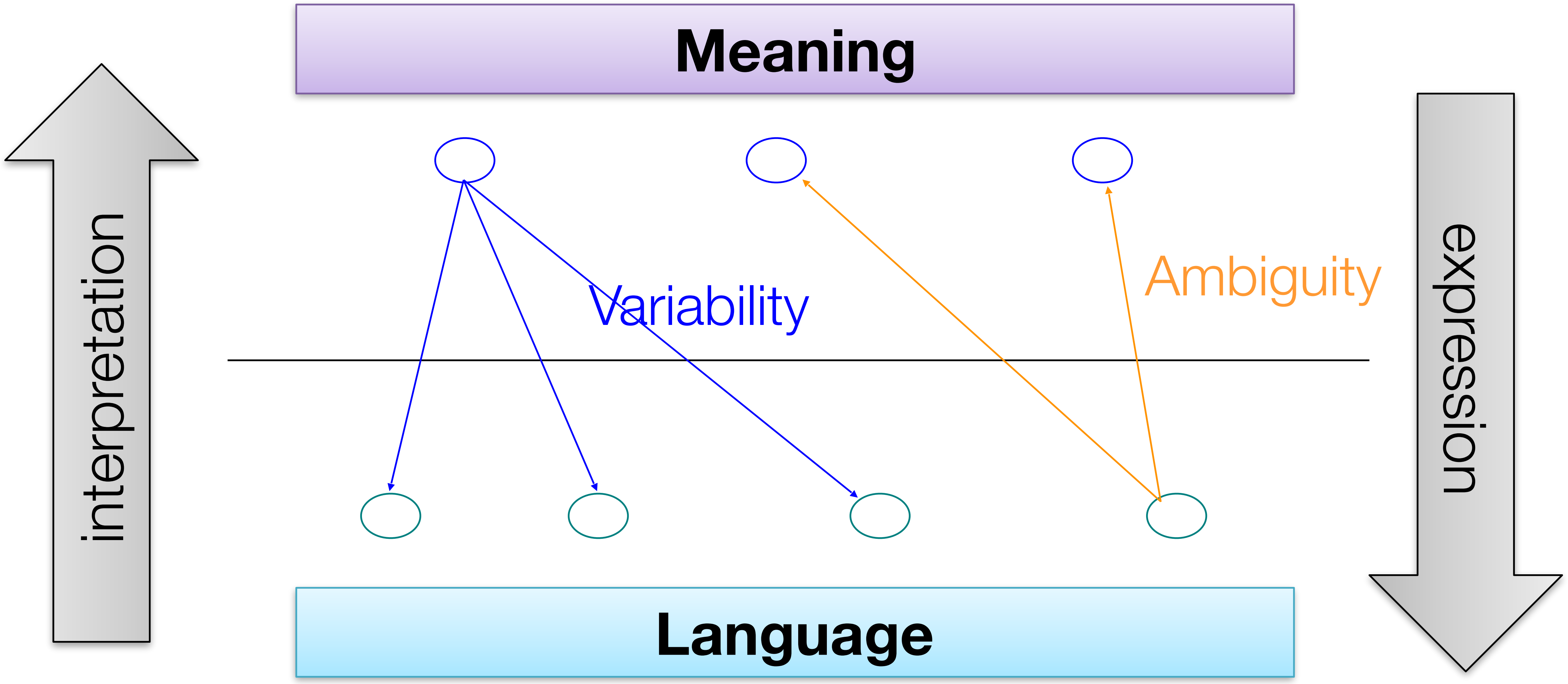
Natural Language and Meaning

Shot an elephant in my pajamas



Natural Language and Meaning





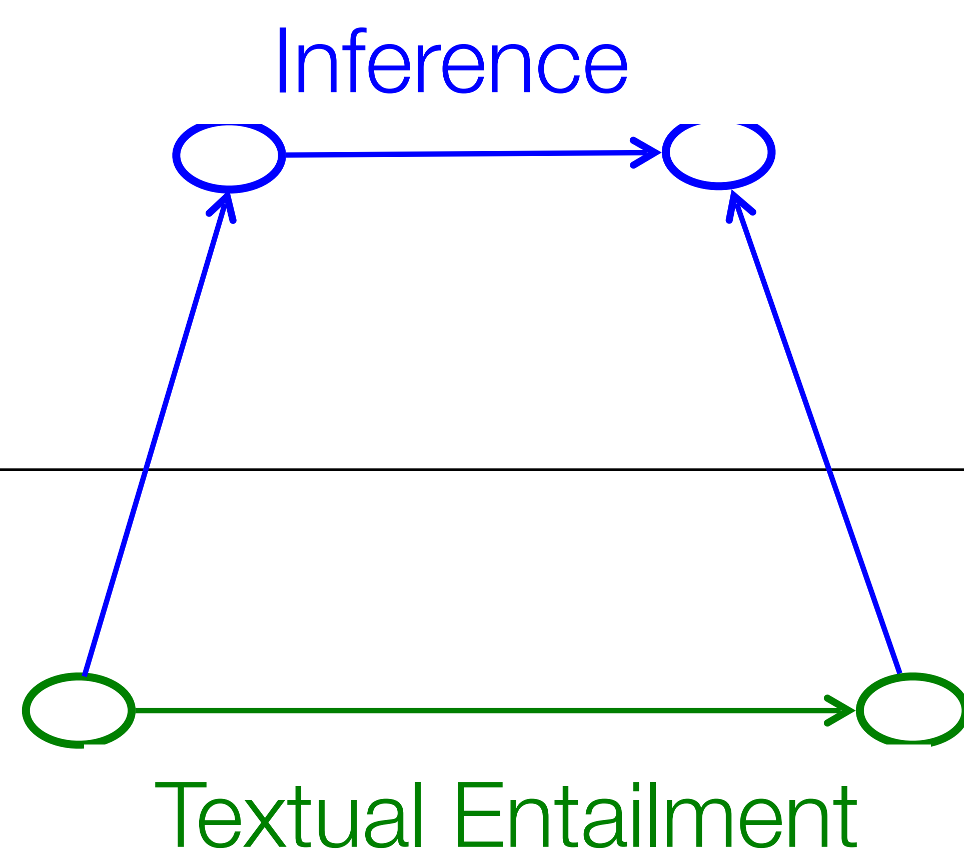
Logical Inference
VS
Textual Entailment

Shot an elephant
in my pajamas

- symbolic – logical forms
- statistical – embeddings

Meaning Representation

Natural Language



Entailment versus Paraphrasing

Entailment versus Paraphrasing

The Dow Jones Industrial Average closed up 255

Entailment versus Paraphrasing

The Dow Jones Industrial Average closed up 255

Dow ends up 255

Dow climbs 255

Dow gains 255 points

Stock market hits a record high

Entailment versus Paraphrasing

The Dow Jones Industrial Average closed up 255

Dow ends up 255

Dow climbs 255

Dow gains 255 points

Stock market hits a record high

Equivalence: text1 ⇔ text2 (paraphrasing)

Entailment versus Paraphrasing

The Dow Jones Industrial Average closed up 255

Dow ends up 255

Dow climbs 255

Dow gains 255 points

Stock market hits a record high

Equivalence: text1 ⇔ text2 (paraphrasing)

Entailment: text1 ⇒ text2

Textual Entailment: Definition

- A directional relation between two text fragments: Text (t) and Hypothesis (h):

***t* entails *h* ($t \Rightarrow h$) if**
humans reading *t* will infer that *h* is most likely true

- Assuming “common background knowledge” – which is indeed expected from applications

Some Examples of Potential Textual Entailment

- T: Legally, John could drive.
- H: John drove.

[Braz et. al. IJCAI workshop’05;PARC Corpus]

Some Examples of Potential Textual Entailment

- T: Legally, John could drive.
- H: John drove.

- T: Bush said that Khan sold centrifuges to North Korea.
- H: Centrifuges were sold to North Korea.

[Braz et. al. IJCAI workshop'05;PARC Corpus]

Some Examples of Potential Textual Entailment

- T: Legally, John could drive.
- H: John drove.

- T: Bush said that Khan sold centrifuges to North Korea.
- H: Centrifuges were sold to North Korea.

- T: No US congressman visited Iraq until the war.
- H: Some US congressmen visited Iraq before the war.

- T: The room was full of men.

[Braz et. al. IJCAI workshop'05;PARC Corpus]

Some Examples of Potential Textual Entailment

- T: Legally, John could drive.
- H: John drove.

- T: Bush said that Khan sold centrifuges to North Korea.
- H: Centrifuges were sold to North Korea.

- T: No US congressman visited Iraq until the war.
- H: Some US congressmen visited Iraq before the war.

- T: The room was full of men.
- H: The room was full of intelligent men.

- T: The New York Times reported that Hanssen sold FBI secrets to the Russians and could face the death penalty.
- H: Hanssen sold FBI secrets to the Russians.

[Braz et. al. IJCAI workshop'05;PARC Corpus]

Some Examples of Potential Textual Entailment

- T: Legally, John could drive.
- H: John drove.

- T: Bush said that Khan sold centrifuges to North Korea.
- H: Centrifuges were sold to North Korea.

- T: No US congressman visited Iraq until the war.
- H: Some US congressmen visited Iraq before the war.

- T: The room was full of men.
- H: The room was full of intelligent men.

- T: The New York Times reported that Hanssen sold FBI secrets to the Russians and could face the death penalty.
- H: Hanssen sold FBI secrets to the Russians.

- T: All soldiers were killed in the ambush.
- H: Many soldiers were killed in the ambush.

[Braz et. al. IJCAI workshop’05;PARC Corpus]

Textual Entailment with ***Knowledge***

***t* entails *h* (*t* ⇒ *h*) if**
humans reading *t* will infer that *h* is most likely true

- For textual entailment to hold we require:
  - *text* AND *knowledge* ⇒ *h*

  but
  - *knowledge* **should not entail** *h* alone
  - Justification: consider time-dependent information, e.g. PresidentOf(US, X)
- Systems are not supposed to validate ***h***'s truth regardless of ***t*** (e.g. by searching ***h*** on the web)

[id: 5T-39 entail]

TEXT: …While no one accuses Madonna of doing anything illegal in adopting the 4-year-old girl, reportedly named Mercy, there are questions nonetheless about how Madonna is able to navigate Malawi's 18-to-24 month vetting period in just a matter of days or weeks…

HYPOTHESIS:

Madonna is 50 years old.

Contradiction: Definition

Contradiction: Definition

- Definition:

  The Hypothesis H of an entailment pair contradicts the Text T if the relations/events described by H are highly unlikely to be true given the relations/events described by T.

- Justification: filtering facts from diverse/noisy sources, detecting state changes

Entailment / Contradiction / Unknown?

- Text:

  The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

- Hyp 1: BMI acquired an American company.

Entailment / Contradiction / Unknown?

- Text:

  The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

- Hyp 1: BMI acquired an American company.

Entailment / Contradiction / Unknown?

- Text:

  The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

- Hyp 1: BMI acquired an American company.

Entailment / Contradiction / Unknown?

- Text:

  The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

- Hyp 1: BMI acquired an American company.

Entailment / Contradiction / Unknown?

- Text:

  The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

- Hyp 1: BMI acquired an American company.

Entailment / Contradiction / Unknown?

- Text:

  The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

- Hyp 1: BMI acquired an American company.

Entailment / Contradiction / Unknown?

- Text:

  The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

- Hyp 2: BMI bought employee-owned LexCorp for $3.4Bn.

Entailment / Contradiction / Unknown?

- Text:

  The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

- Hyp 2: BMI bought employee-owned LexCorp for $3.4Bn.

Entailment / Contradiction / Unknown?

- Text:

  The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

- Hyp 2: BMI bought employee-owned LexCorp for $3.4Bn.

Entailment / Contradiction / Unknown?

- Text:

  The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

- Hyp 2: BMI bought employee-owned LexCorp for $3.4Bn.

Entailment / Contradiction / Unknown?

- Text:

  The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

- Hyp 2: BMI bought employee-owned LexCorp for $3.4Bn.

Entailment / Contradiction / Unknown?

- Text:

  The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

- Hyp 3: BMI is an employee-owned concern.

Entailment / Contradiction / Unknown?

- Text:

  The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

- Hyp 3: BMI is an employee-owned concern.

Entailment / Contradiction / Unknown?

- Text:

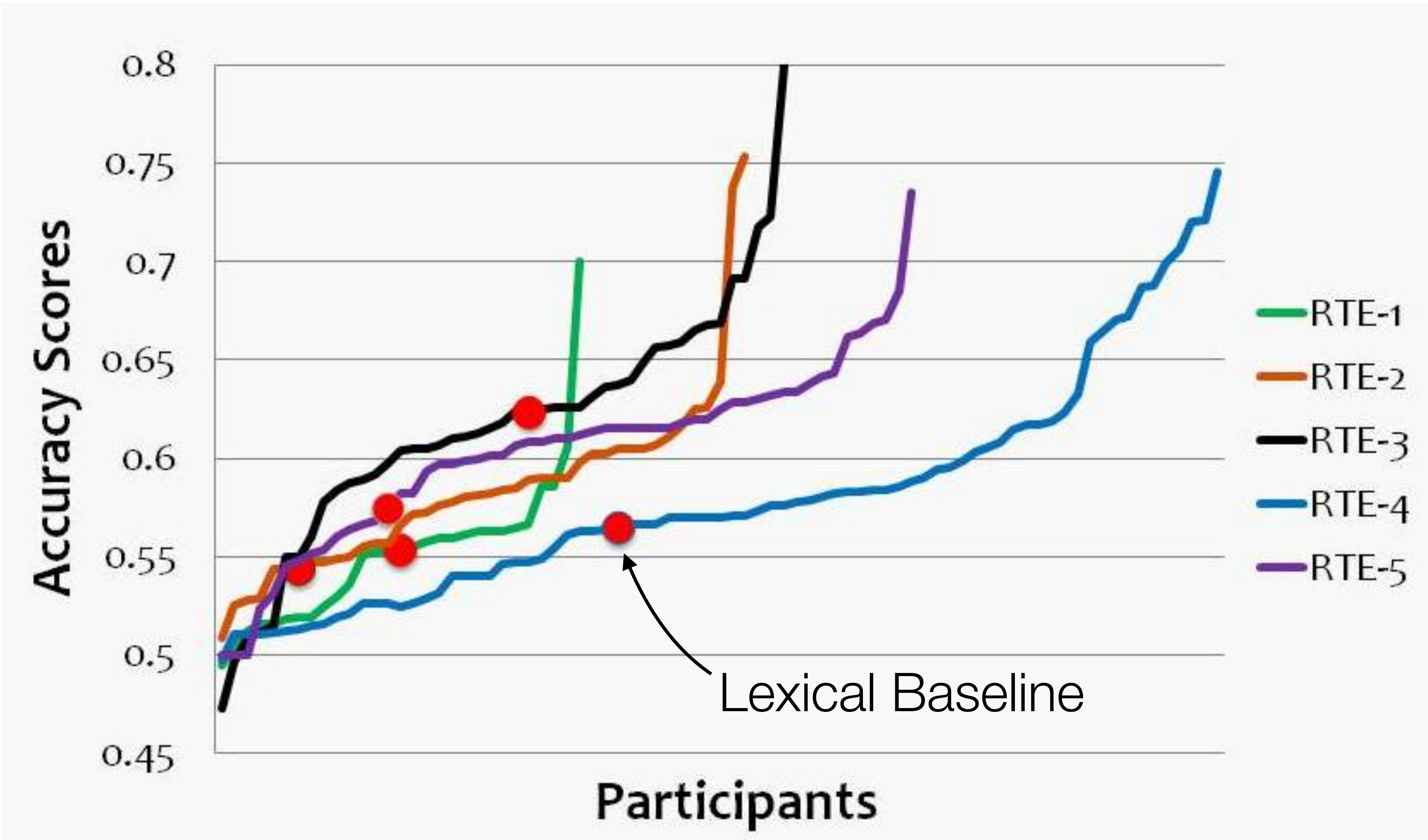
  The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

- Hyp 3: BMI is an employee-owned concern.

RTE Evaluation

- Examples drawn from NLP tasks/domains
- **~90% pairwise inter-annotator agreement**
- RTE 1-3: ~800 dev, 800 test RTE pairs each (‘05- ‘07)
  - Boolean label: “entailed” vs. “not entailed”
  - BALANCED data set
- RTE 4-5: Ave. text length = 40,100 words (‘08, ‘09) respectively, 2-way and 3-way tasks
  - “entailed”, “contradicted”, and “unknown”
- Some pilot RTE task data sets as well
- RTE 6 (2010): shift to application focus: IR-like setting

How well did NLP do?




Will
Work
For FOOd

Alignment for Entailment

Alignment for RTE

- Idea: break entailment into smaller decisions
- Alignment as a way to recognize relevant Text portions
- Portions of text compared using closed set of operations
  - Operations include lexical similarity, structural similarity
  - Possible to define concepts such as semantic containment and semantic exclusion
  - May be extended using Knowledge bases

Alignment for RTE

- Impose ***constraints*** on the aggregate set of comparisons we entertain
- E.g. each Hypothesis element can match at most one Text element

Alignment for RTE

- Impose ***constraints*** on the aggregate set of comparisons we entertain

- E.g. each Hypothesis element can match at most one Text element

Alignment: a mapping from elements in the Hypothesis to elements in the Text under specified constraints

Alignment for RTE

- Impose ***constraints*** on the aggregate set of comparisons we entertain

- E.g. each Hypothesis element can match at most one Text element

Alignment: a mapping from elements in the Hypothesis to elements in the Text under specified constraints

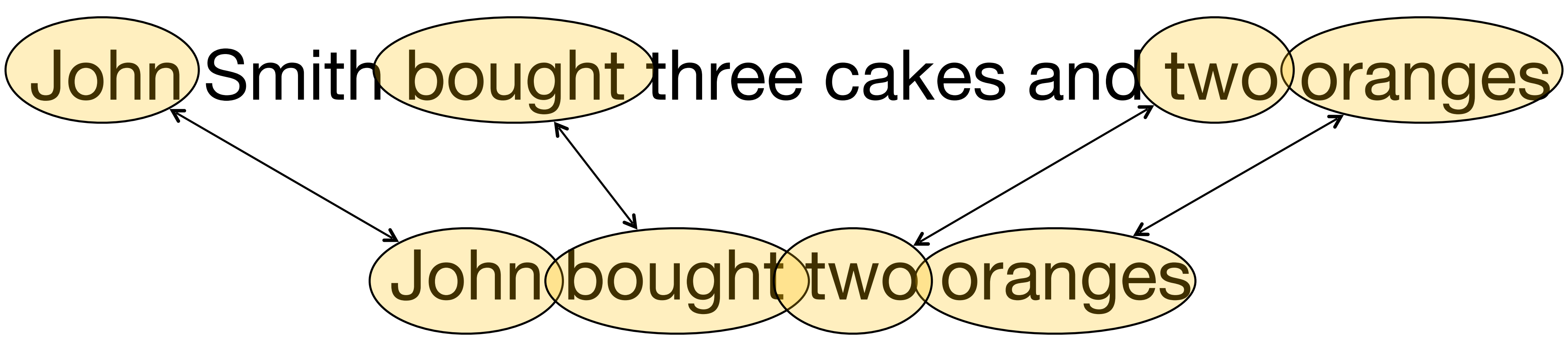
John Smith bought three cakes and two oranges

John bought two oranges

Alignment for RTE

- Impose ***constraints*** on the aggregate set of comparisons we entertain
- E.g. each Hypothesis element can match at most one Text element

Alignment: a mapping from elements in the Hypothesis to elements in the Text under specified constraints



Shallow Alignment as Focus Of Attention

- Pick a “good” shallow alignment
- Use this to query deeper structure/extract features

Shallow Alignment as Focus Of Attention

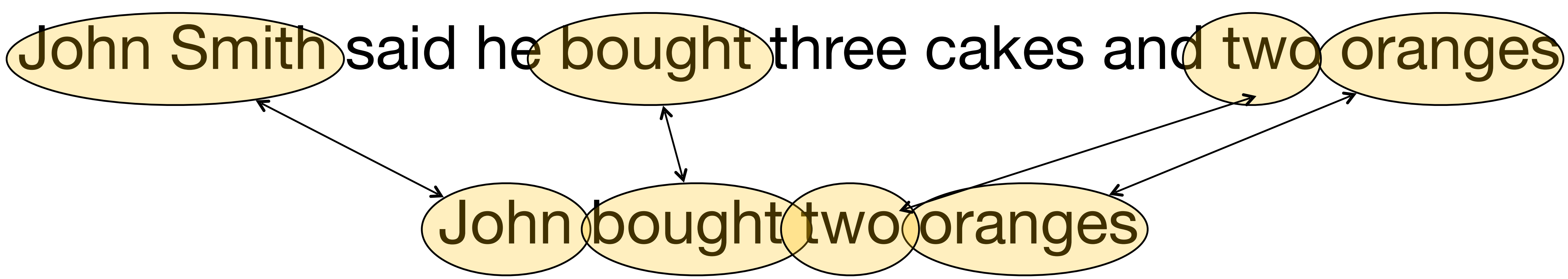
- Pick a “good” shallow alignment
- Use this to query deeper structure/extract features

John Smith said he bought three cakes and two oranges

John bought two oranges

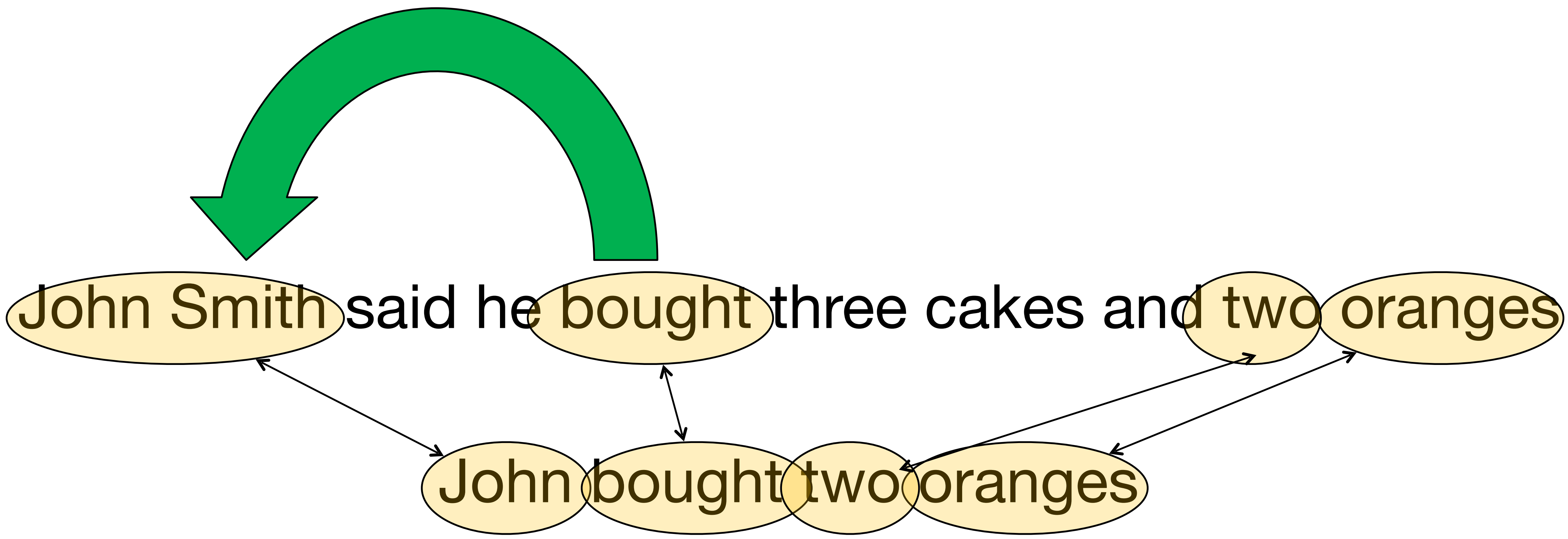
Shallow Alignment as Focus Of Attention

- Pick a “good” shallow alignment
- Use this to query deeper structure/extract features



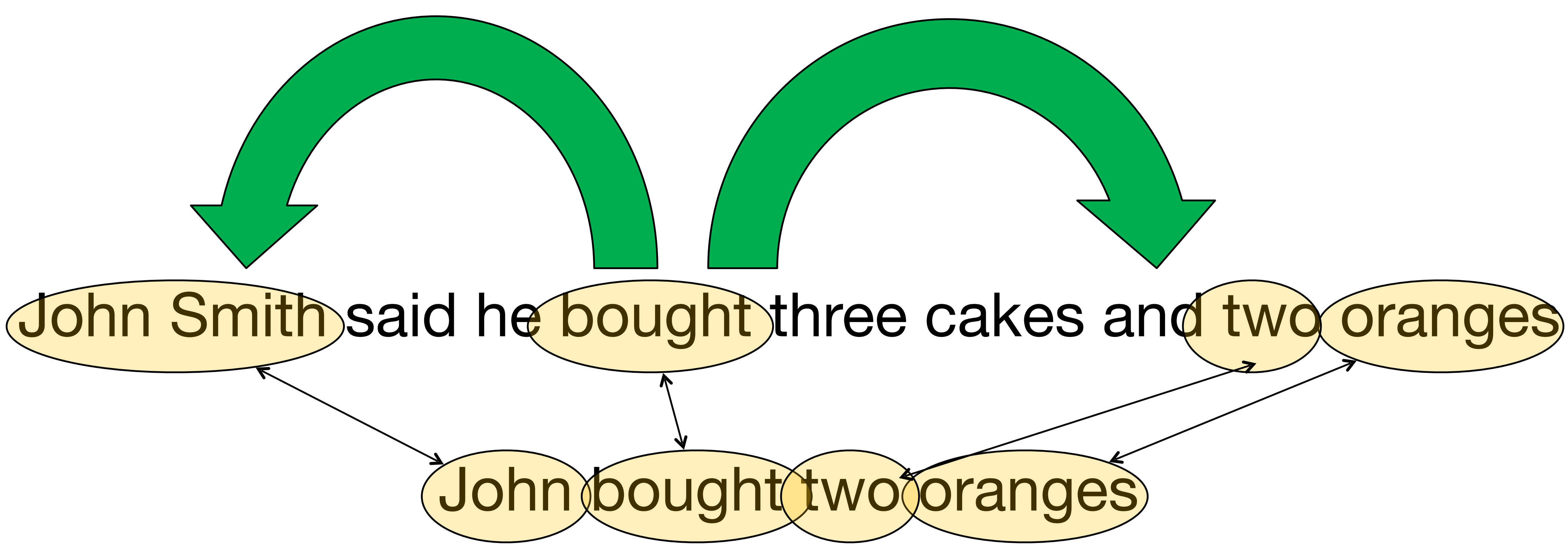
Shallow Alignment as Focus Of Attention

- Pick a “good” shallow alignment
- Use this to query deeper structure/extract features



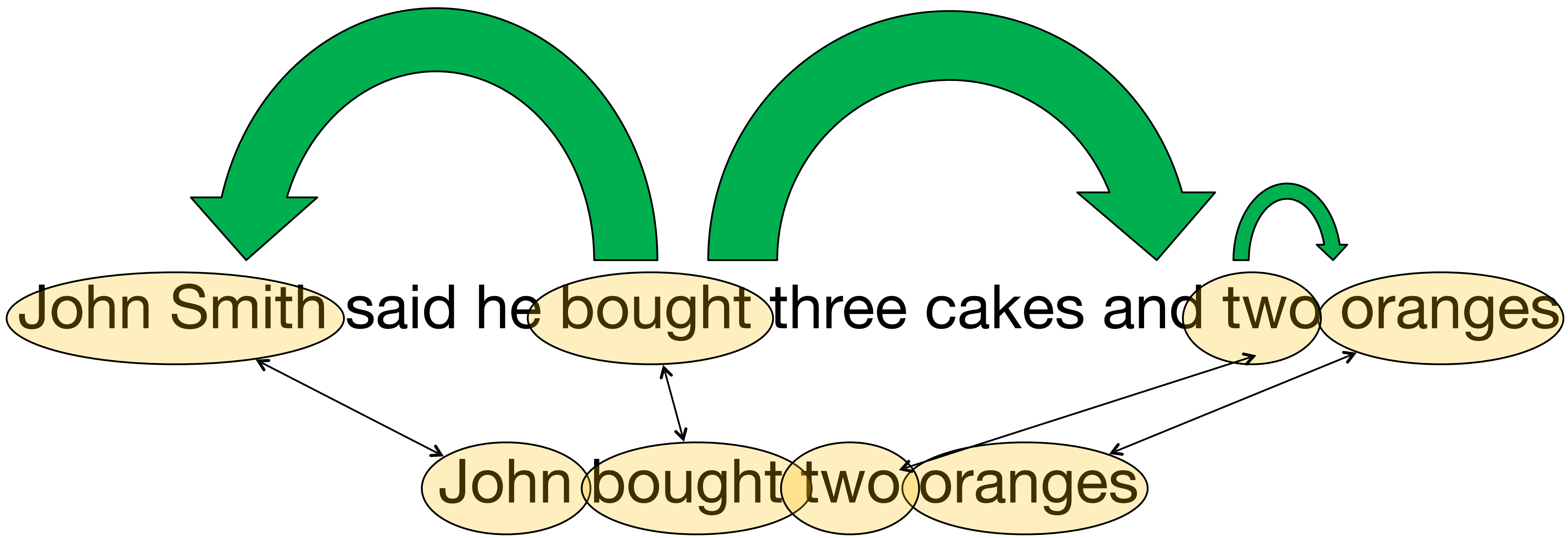
Shallow Alignment as Focus Of Attention

- Pick a “good” shallow alignment
- Use this to query deeper structure/extract features



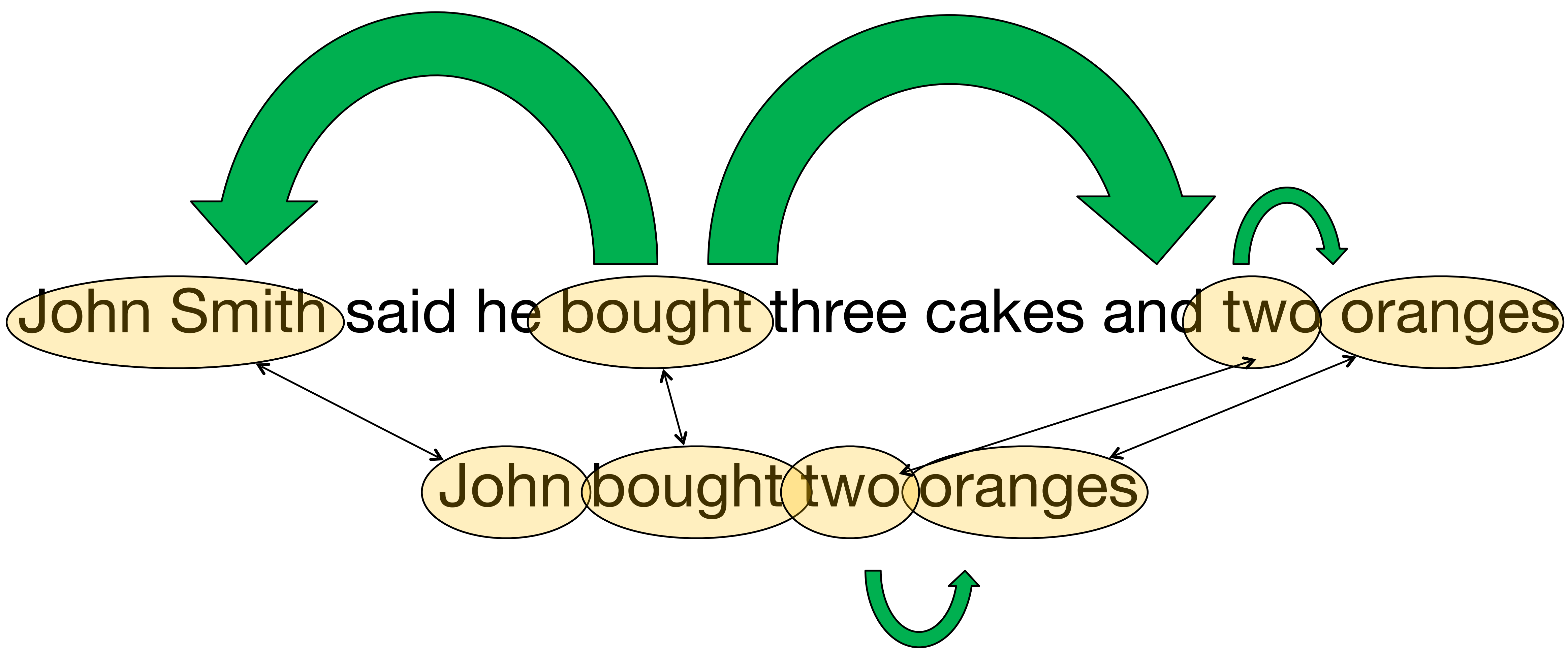
Shallow Alignment as Focus Of Attention

- Pick a “good” shallow alignment
- Use this to query deeper structure/extract features



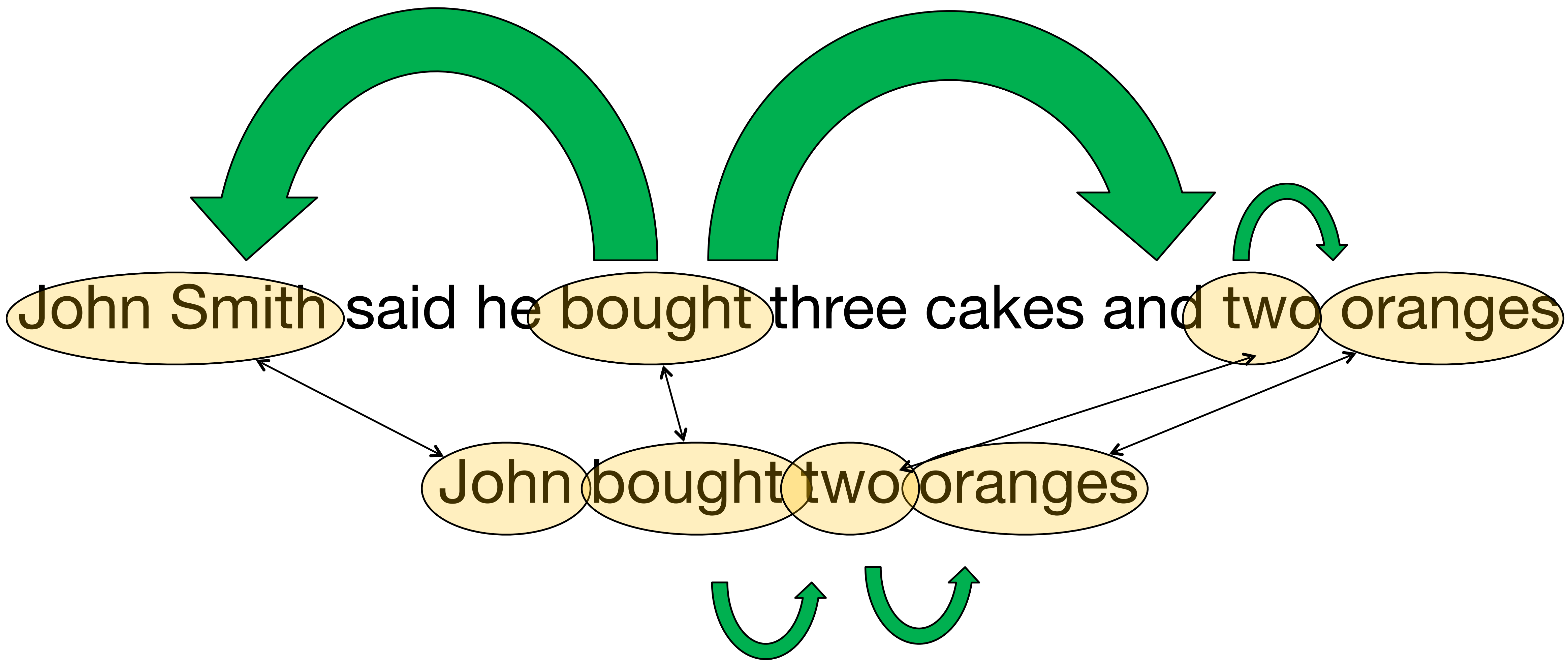
Shallow Alignment as Focus Of Attention

- Pick a “good” shallow alignment
- Use this to query deeper structure/extract features



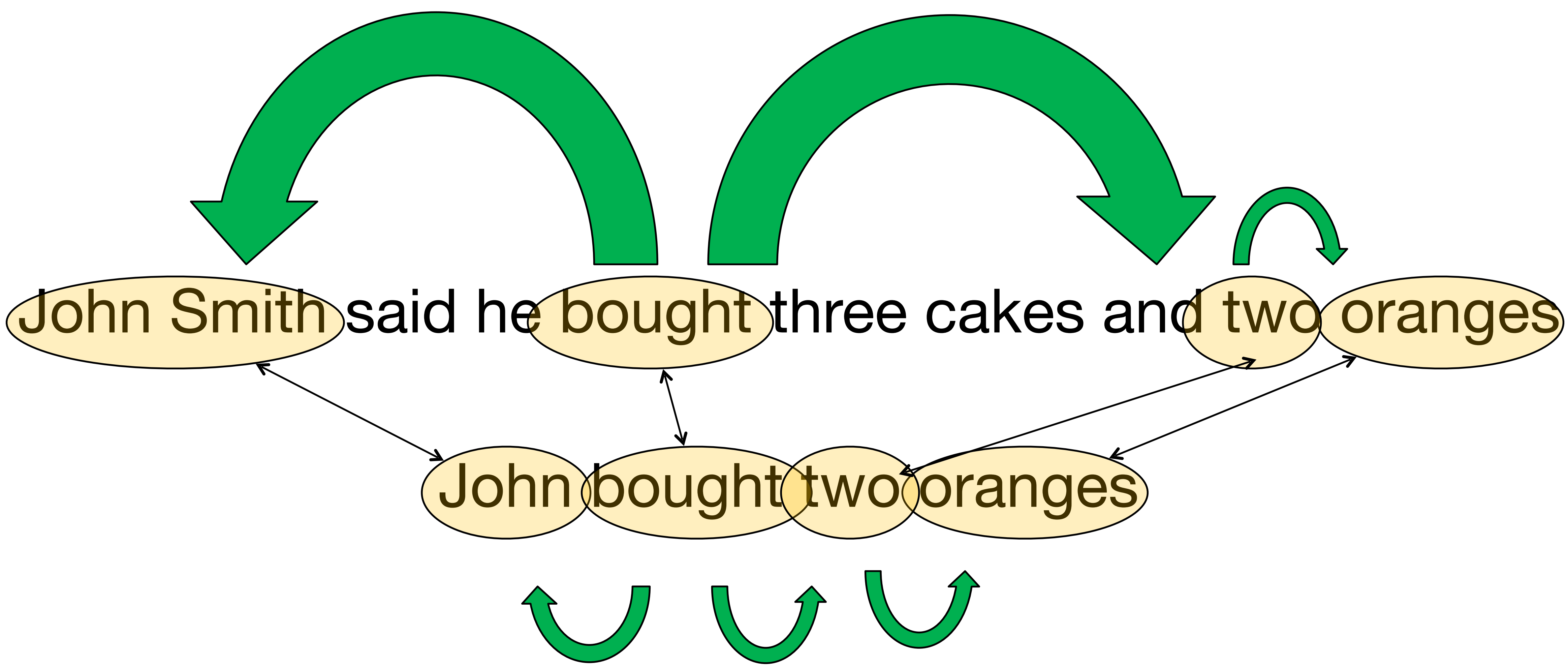
Shallow Alignment as Focus Of Attention

- Pick a “good” shallow alignment
- Use this to query deeper structure/extract features



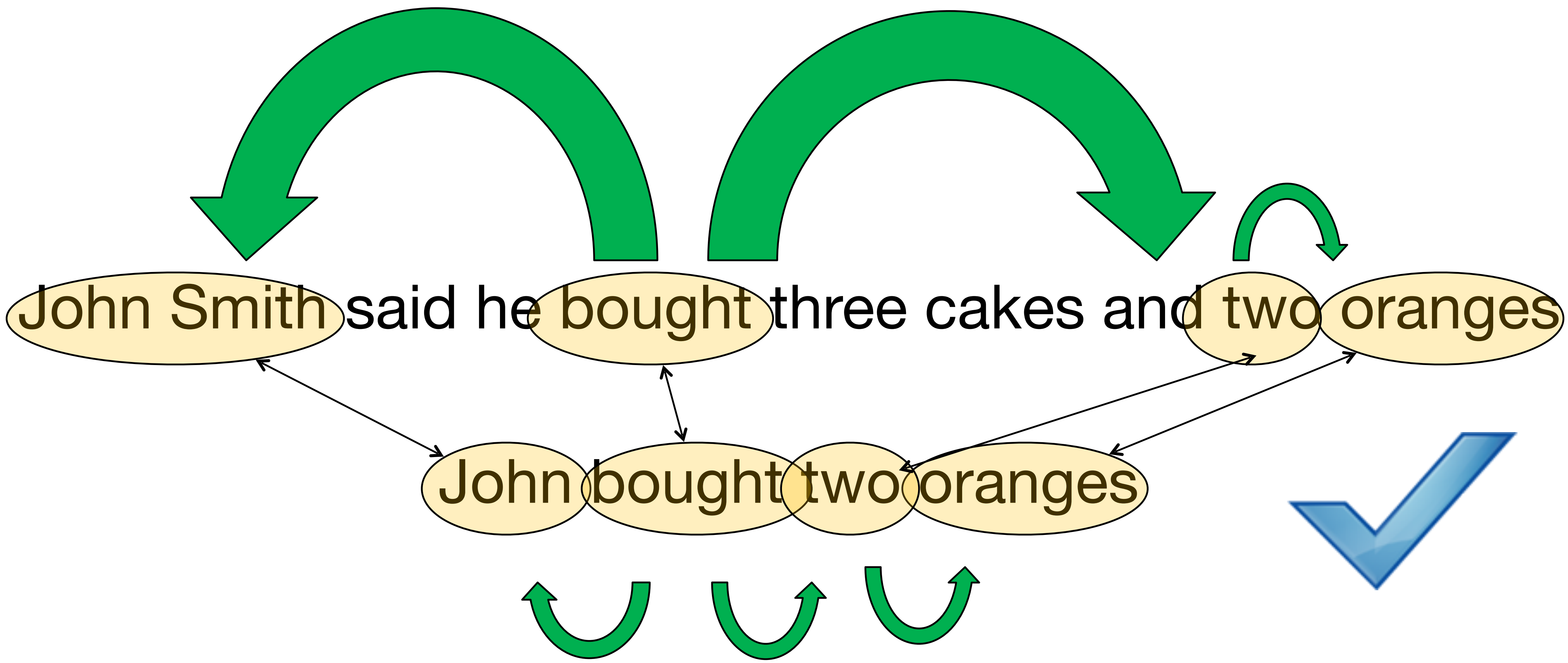
Shallow Alignment as Focus Of Attention

- Pick a "good" shallow alignment
- Use this to query deeper structure/extract features



Shallow Alignment as Focus Of Attention

- Pick a “good” shallow alignment
- Use this to query deeper structure/extract features



Shallow Alignment as Focus Of Attention

- Pick a “good” shallow alignment
- Use this to query deeper structure/extract features

Shallow Alignment as Focus Of Attention

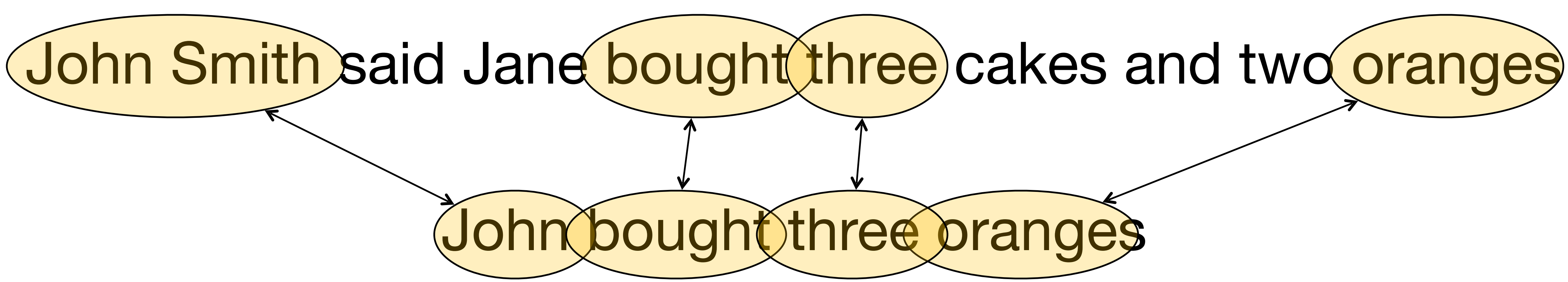
- Pick a “good” shallow alignment
- Use this to query deeper structure/extract features

John Smith said Jane bought three cakes and two oranges

John bought three oranges

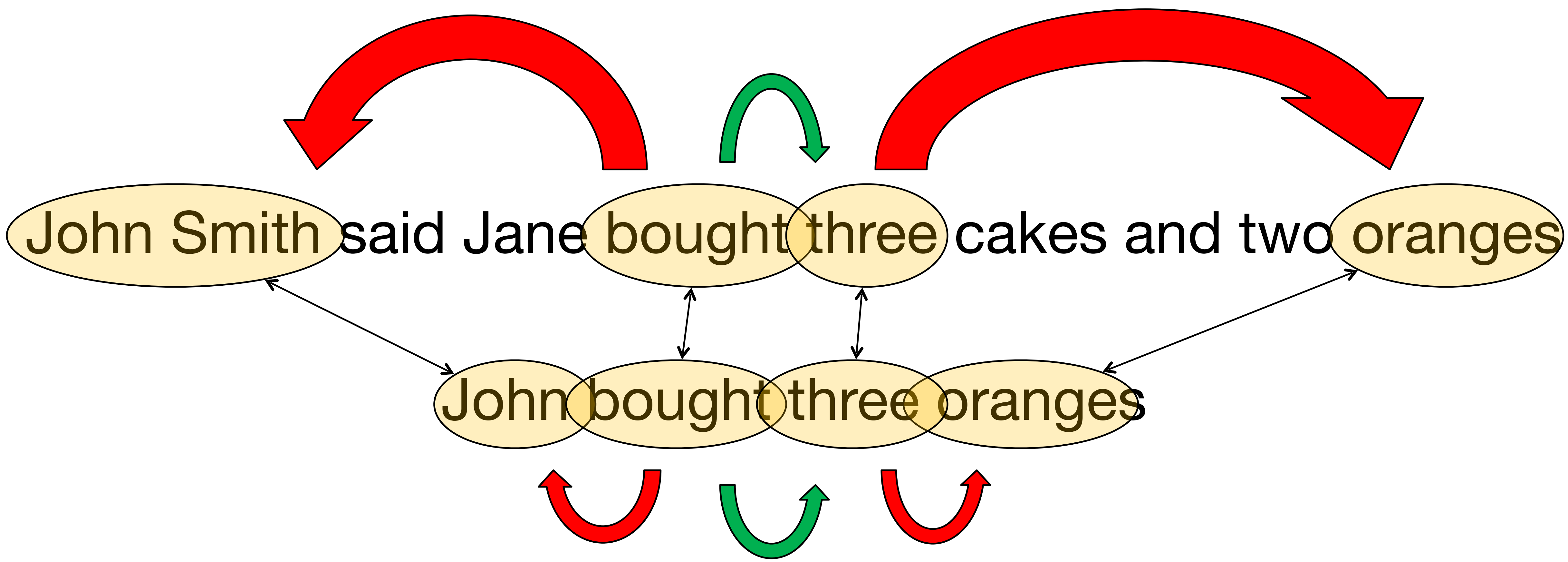
Shallow Alignment as Focus Of Attention

- Pick a “good” shallow alignment
- Use this to query deeper structure/extract features



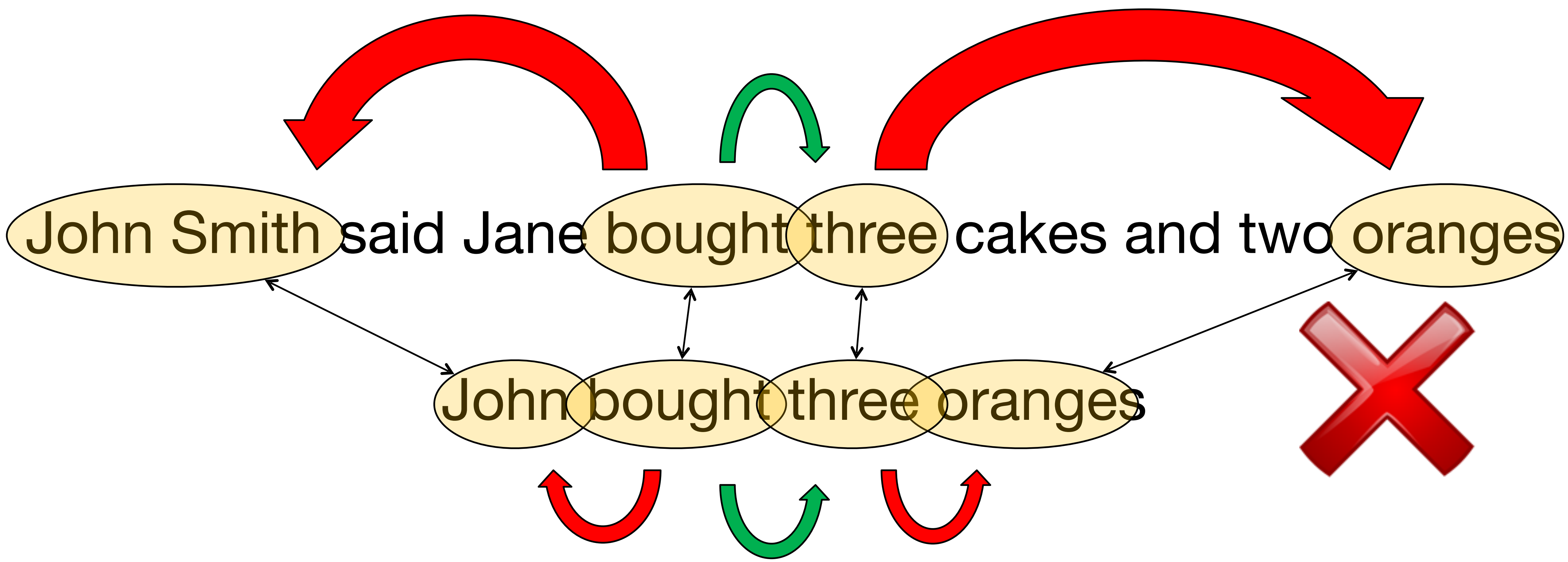
Shallow Alignment as Focus Of Attention

- Pick a "good" shallow alignment
- Use this to query deeper structure/extract features



Shallow Alignment as Focus Of Attention

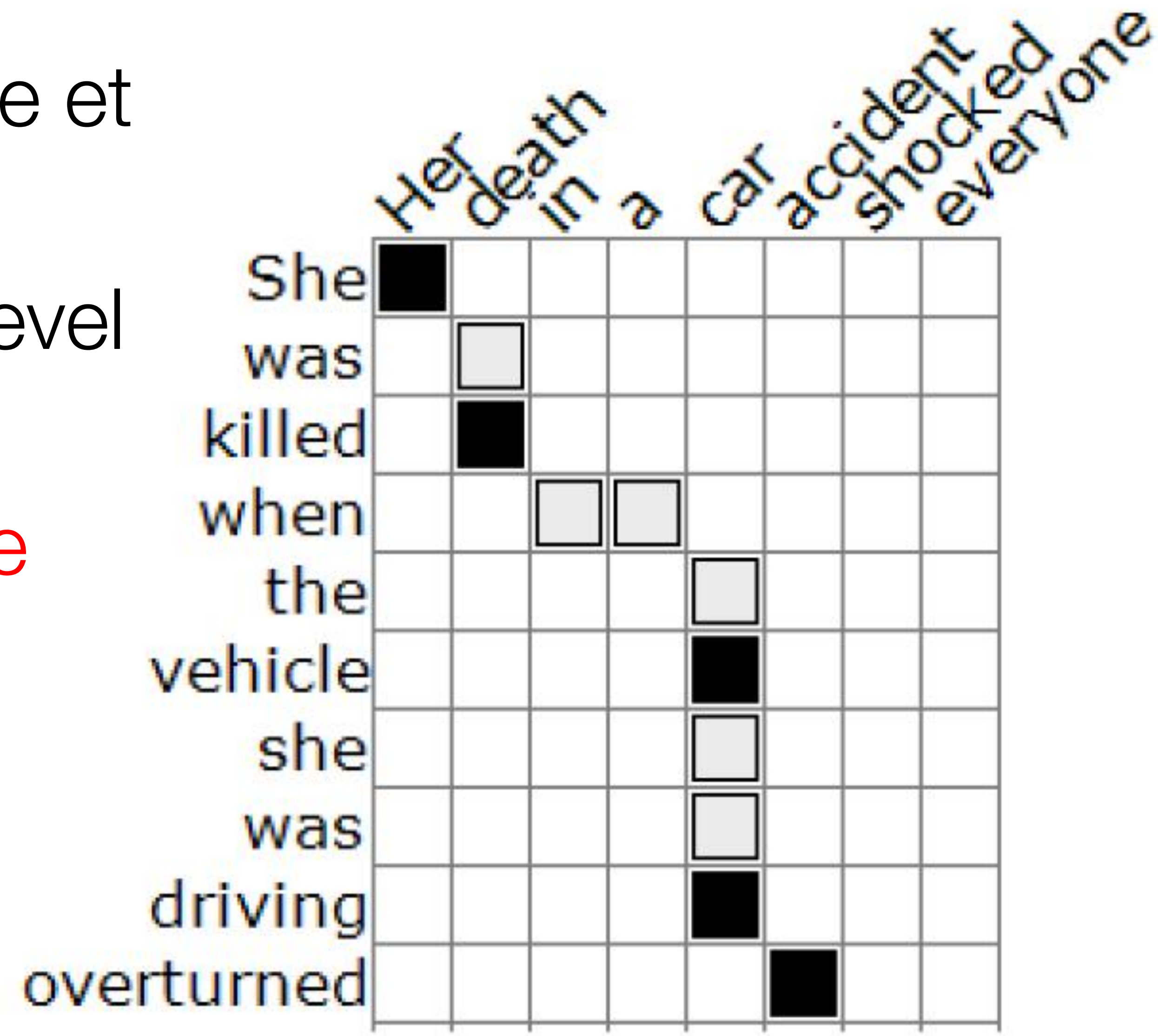
- Pick a “good” shallow alignment
- Use this to query deeper structure/extract features



Alignment for RTE

Chambers et al. 2007, deMarneffe et al. 2007

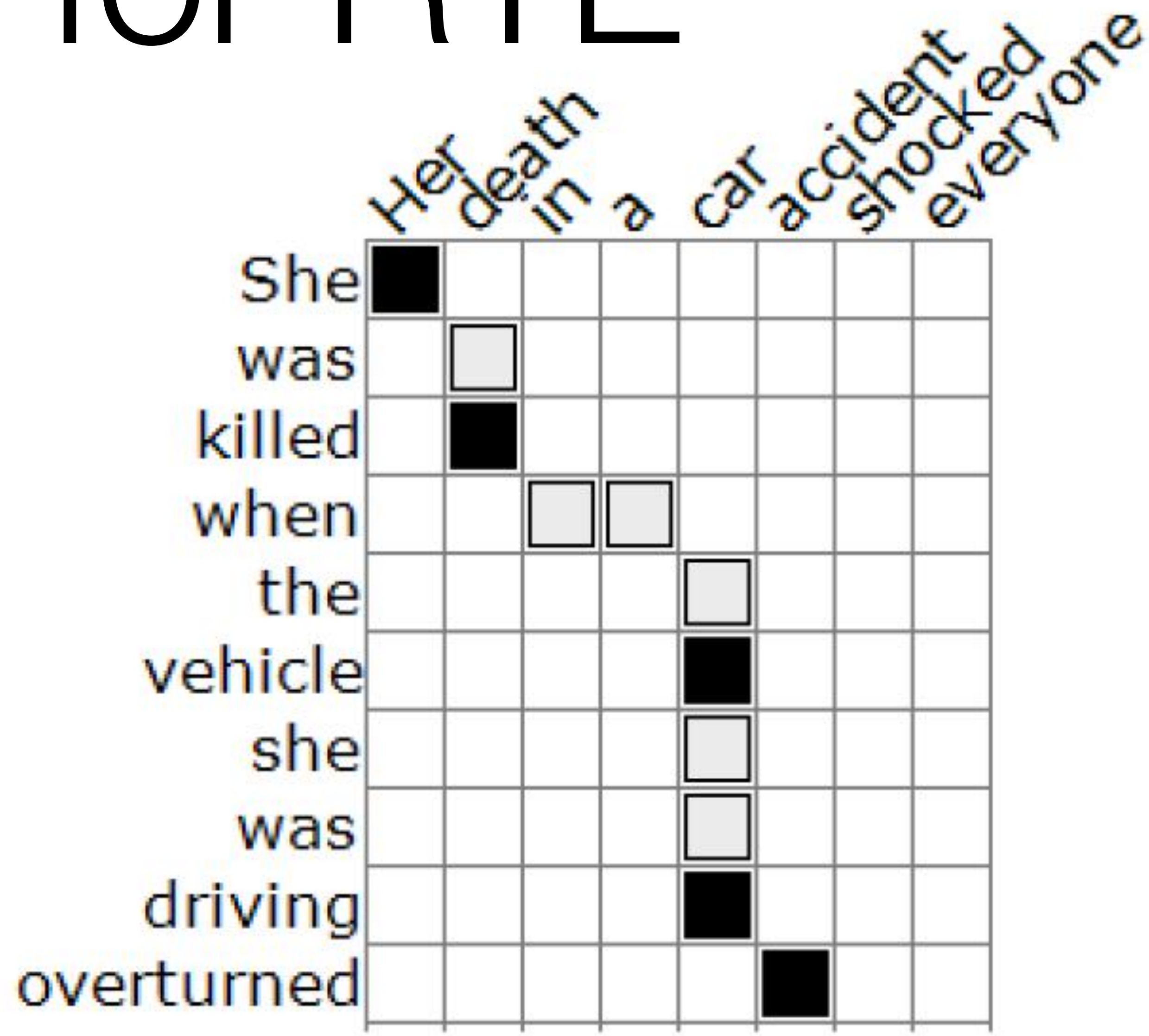
- learn "alignment" from lexical-level labelings
    - Intuition: abstract away some logical structure, irrelevant content
    - Identify the parts of T that "support" H
- Identify "relevant" parts of T via word, edge weight vectors



Alignment for RTE

Chambers et al. 2007,
deMarneffe et al. 2007

- Use alignment to extract features for discerning "entailed" from "not entailed", using deeper semantic structure



$$\text{score}(a) = \sum_{i \in h} \text{score}_w(h_i, a(h_i)) + \sum_{(i,j) \in e(h)} \text{score}_e((h_i, h_j), (a(h_i), a(h_j)))$$

Why does alignment work? (when it does...)

Why does alignment work? (when it does...)

- Comparable to similarity metric approach
  - Trying to capture deeper structure

Why does alignment work? (when it does...)

- Comparable to similarity metric approach
  - Trying to capture deeper structure
- Supports discriminative ML by generating sufficiently coarse features

Why does alignment work? (when it does...)

- Comparable to similarity metric approach
  - Trying to capture deeper structure
- Supports discriminative ML by generating sufficiently coarse features
- Works best on cases where content in H is explicit in T
  - But with better deep structure/appropriate representation, expect to do better

Why does alignment work? (when it does...)

- Comparable to similarity metric approach
  - Trying to capture deeper structure
- Supports discriminative ML by generating sufficiently coarse features
- Works best on cases where content in H is explicit in T
  - But with better deep structure/appropriate representation, expect to do better
- Better inputs => better alignments
  - Problem: pipeline effect for erroneous annotations AND for erroneous alignment

Problems with alignment

Problems with alignment

- Mapping “relevant” parts may be correct intuition, but “relevant” seems to depend on deep structure
  - Fixed heuristic/learned mapping based on shallow cues is problematic
  - Distance is not a reliable proxy for deep structure

Problems with alignment

- Mapping "relevant" parts may be correct intuition, but "relevant" seems to depend on deep structure
  - Fixed heuristic/learned mapping based on shallow cues is problematic
  - Distance is not a reliable proxy for deep structure
- May be multiple match candidates for many H constituent (i.e., shallow alignment may pick the wrong one)
  - Alignment constraints introduce a problem in fixed two-stage system

Alternative: Using Structure as Focus Of Attention

- Find best structural match
- Base entailment results on results of shallow comparison resources

Alternative: Using Structure as Focus Of Attention

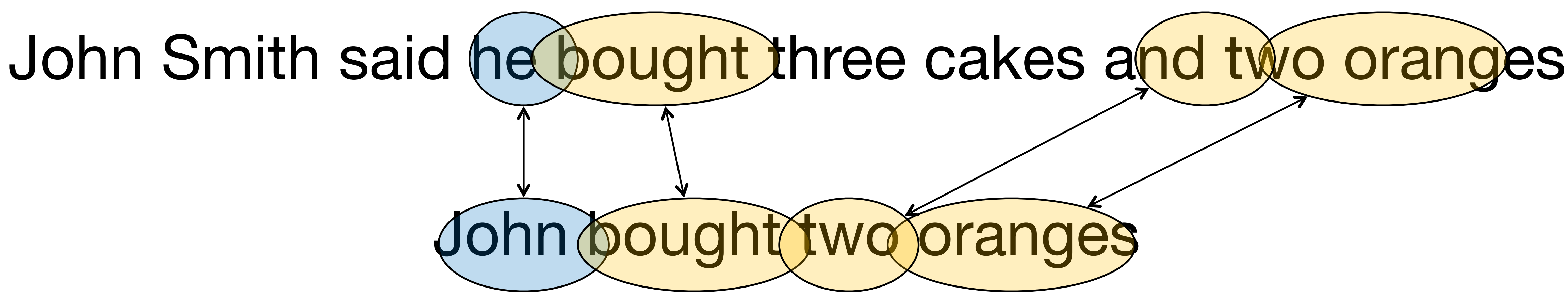
- Find best structural match
- Base entailment results on results of shallow comparison resources

John Smith said he bought three cakes and two oranges

John bought two oranges

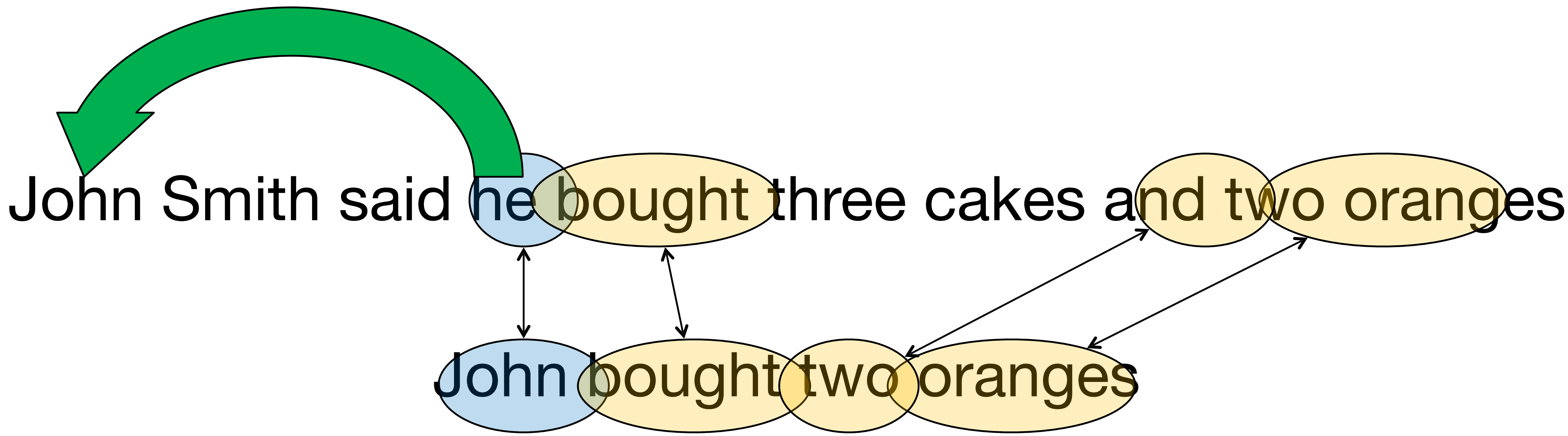
Alternative: Using Structure as Focus Of Attention

- Find best structural match
- Base entailment results on results of shallow comparison resources



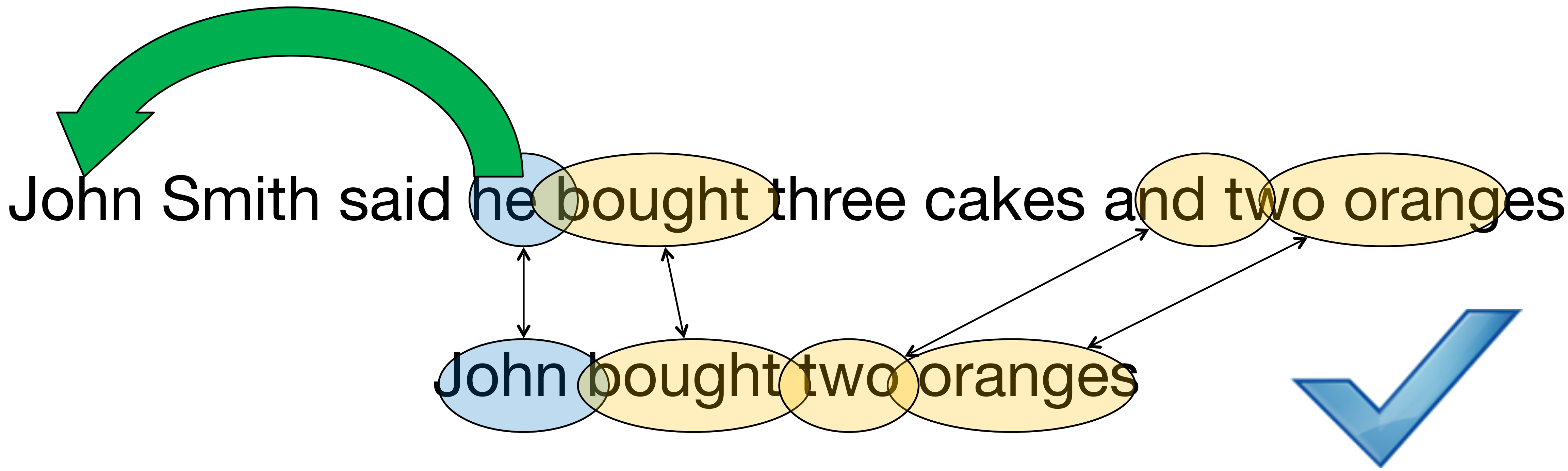
Alternative: Using Structure as Focus Of Attention

- Find best structural match
- Base entailment results on results of shallow comparison resources



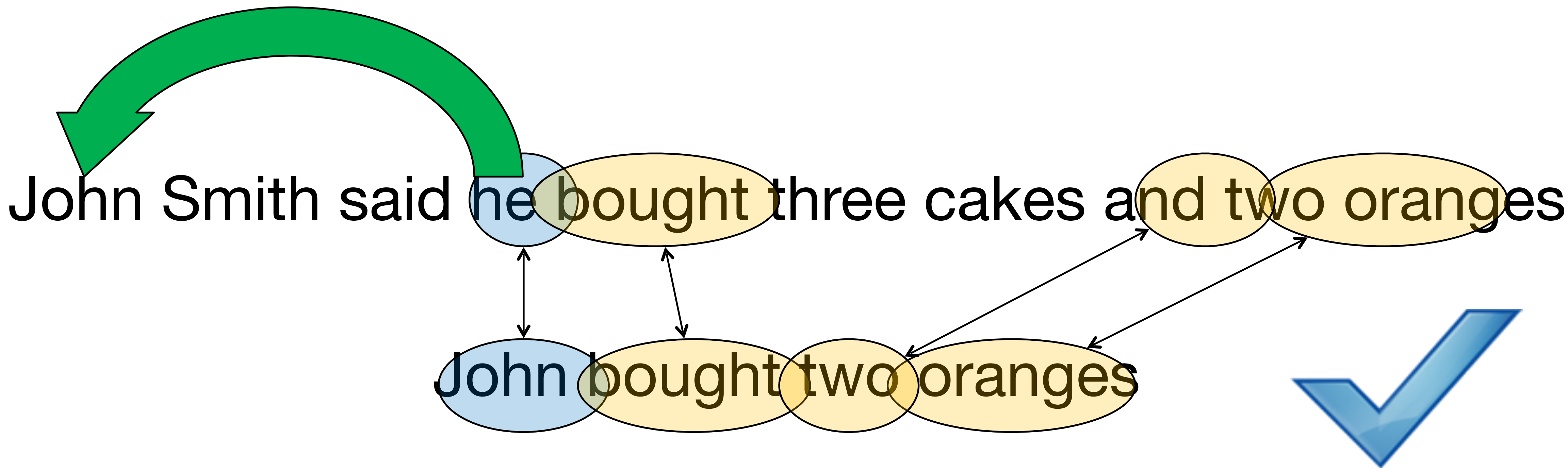
Alternative: Using Structure as Focus Of Attention

- Find best structural match
- Base entailment results on results of shallow comparison resources



Alternative: Using Structure as Focus Of Attention

- Find best structural match
- Base entailment results on results of shallow comparison resources

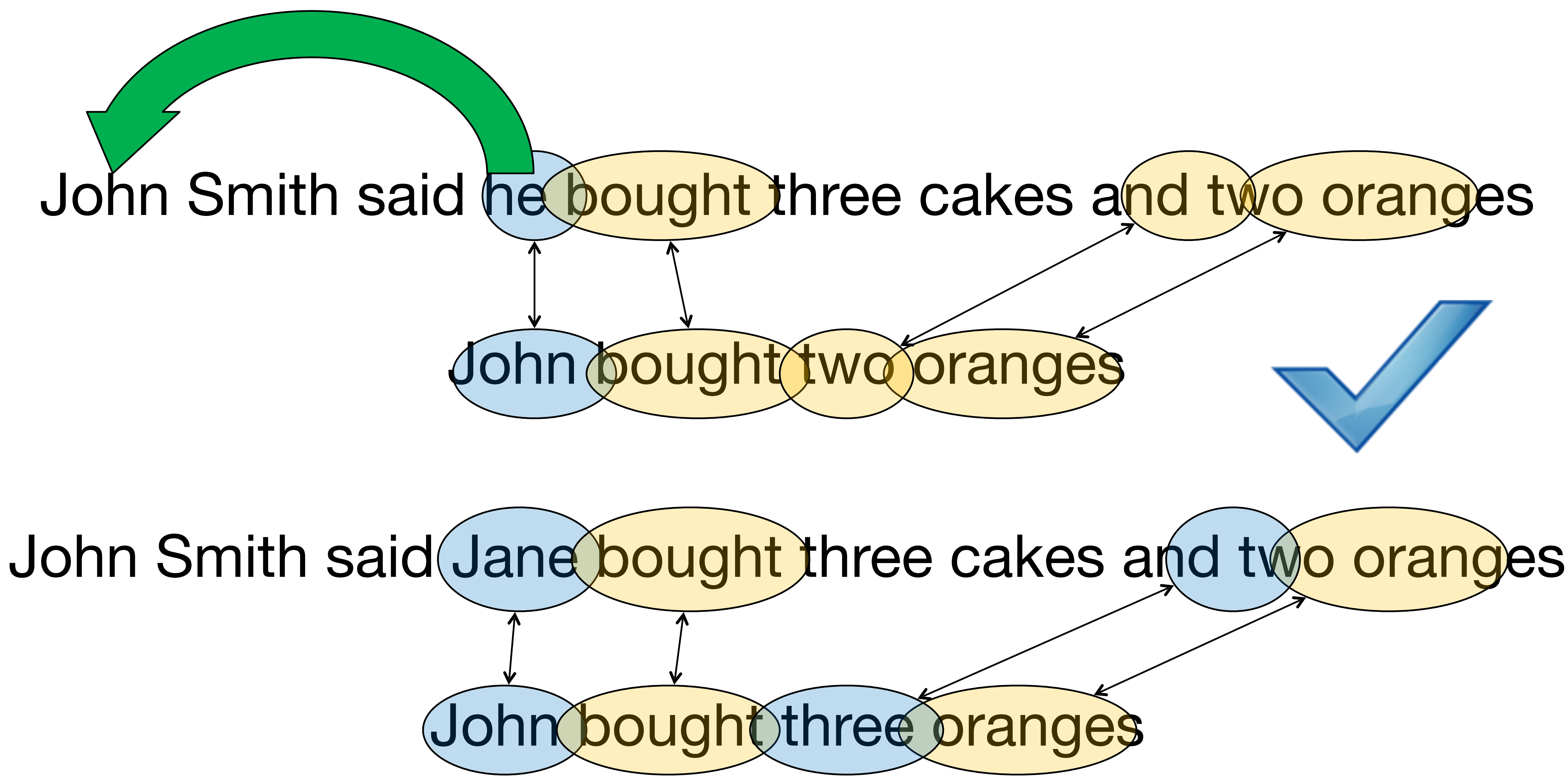


John Smith said Jane bought three cakes and two oranges

John bought three oranges

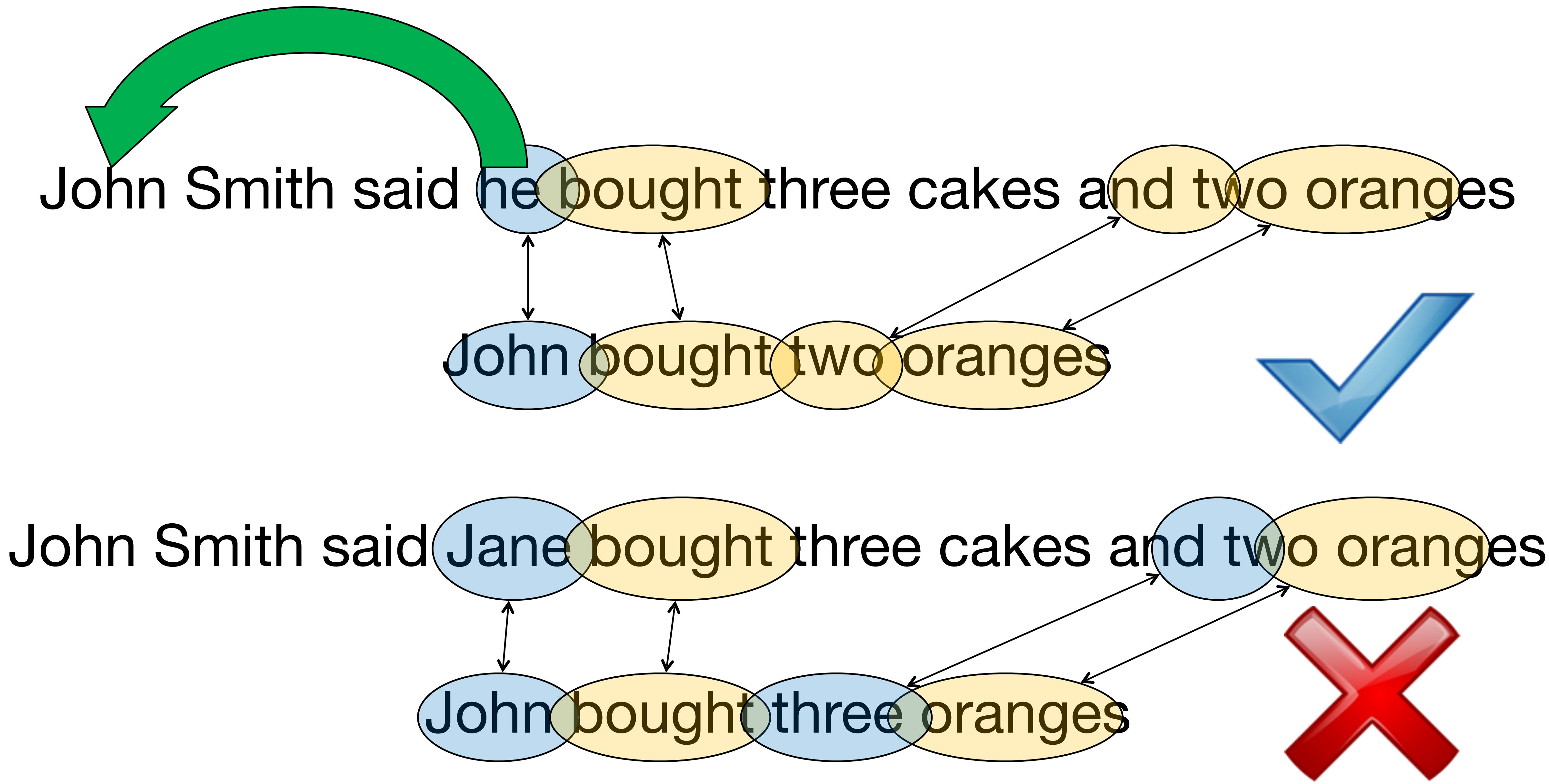
Alternative: Using Structure as Focus Of Attention

- Find best structural match
- Base entailment results on results of shallow comparison resources



Alternative: Using Structure as Focus Of Attention

- Find best structural match
- Base entailment results on results of shallow comparison resources

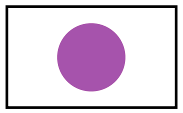
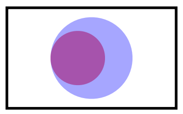
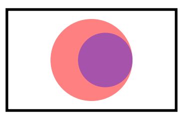
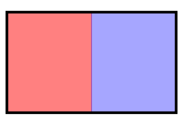
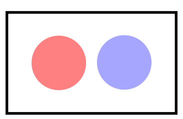
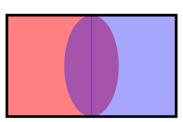
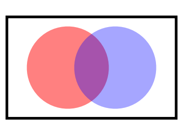


Natural Logic Inference (Natural-LI)

Natural logic (NatLog)

- Use natural logic representation for TE
- Initial implementation of alignment based entailment inference
- Inference patterns built over shallow surface forms, instead of full semantic interpretation

7 basic entailment relations

Venn	symbol	name	example
	P = Q	equivalence	*couch* = *sofa*
	P ⊏ Q	forward entailment (strict)	*crow* ⊏ *bird*
	P ⊐ Q	reverse entailment (strict)	*European* ⊐ *French*
	P ^ Q	negation (exhaustive exclusion)	*human* ^ *nonhuman*
	P | Q	alternation (non-exhaustive exclusion)	*cat* | *dog*
	P _ Q	cover (exhaustive non-exclusion)	*animal* _ *nonhuman*
	P # Q	independence	*hungry* # *hippo*

Relations are defined for all semantic types: *tiny* ⊏ *small*, *hover* ⊏ *fly*, *kick* ⊏ *strike*, *this morning* ⊏ *today*, *in Beijing* ⊏ *in China*, *everyone* ⊏ *someone*, *all* ⊏ *most* ⊏ *some*

Slides based out of Bill MacCartney and Chris Manning's talk in COLING 2008.

Entailment & semantic composition

- Ordinarily, semantic composition preserves entailment relations:

$pork \sqsubset meat$ => $eat\ pork \sqsubset eat\ meat$

$bird \mid fish$ => $big\ bird \mid big\ fish$

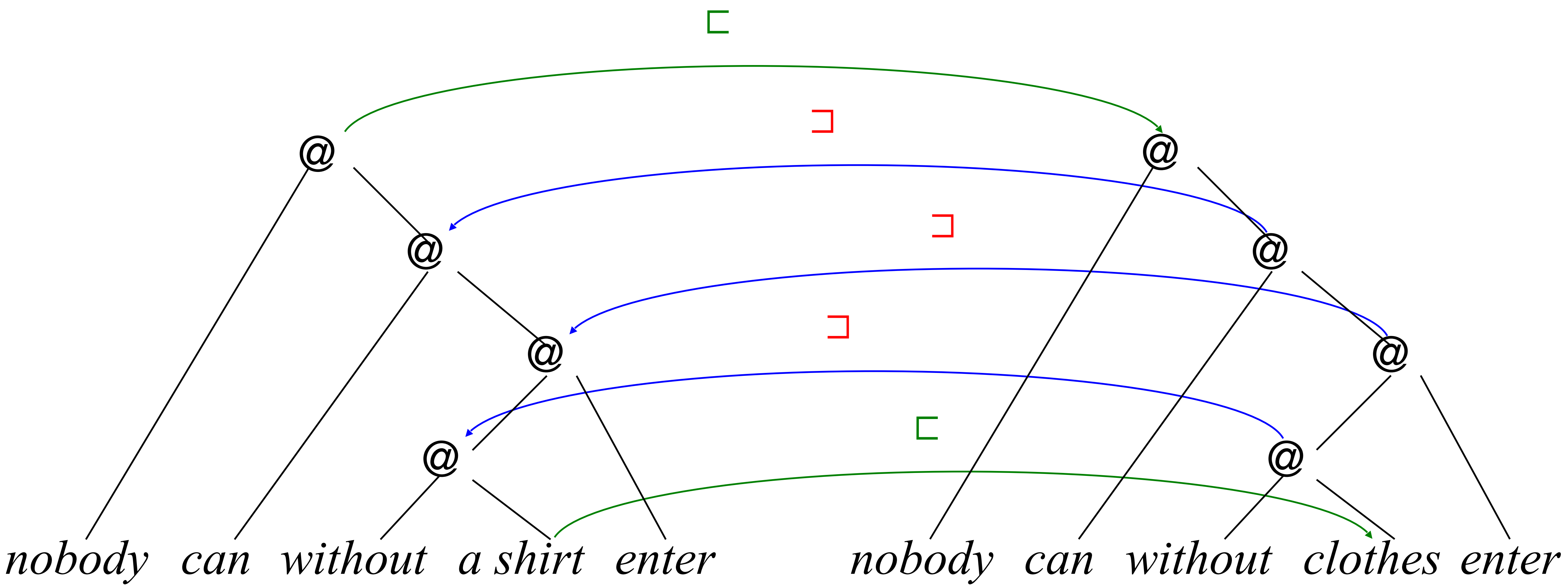
- But many semantic functions behave differently:

$tango \sqsubset dance \quad \Rightarrow \quad refuse\ to\ tango \sqsupset refuse\ to\ dance$

$French \mid German \quad \Rightarrow \quad not\ French \smile not\ German$

Projecting entailment relations upward

- Assume idealized semantic composition trees
- Propagate entailment relation between atoms upward, according to projectivity class of each node on path to root



A (weak) inference procedure

1. Find sequence of edits connecting P and H
   - Insertions, deletions, substitutions, …
2. Determine lexical entailment relation for each edit
   - Substitutions: depends on meaning of substituends: **cat | dog**
   - Deletions: ⊏ by default: **red socks ⊏ socks**
   - But some deletions are special: **not ill ^ ill, refuse to go | go**
   - Insertions are symmetric to deletions: **⊐ by default**
3. Project up to find entailment relation across each edit
4. Compose entailment relations across sequence of edits
   - à la Tarski’s relation algebra

Natural Language Inference

Skip the logical form and do inference from the text directly

Text	Judgments	Hypothesis

https://nlp.stanford.edu/projects/snli/

Skip the logical form and do inference from the text directly

Text	Judgments	Hypothesis
A man inspects the uniform of a figure in some East Asian country.	contradiction C C C C C	The man is sleeping

https://nlp.stanford.edu/projects/snli/

Skip the logical form and do inference from the text directly

Text	Judgments	Hypothesis
A man inspects the uniform of a figure in some East Asian country.	contradiction C C C C C	The man is sleeping
An older and younger man smiling.	neutral N N E N N	Two men are smiling and laughing at the cats playing on the floor.

https://nlp.stanford.edu/projects/snli/

Skip the logical form and do inference from the text directly

Text	Judgments	Hypothesis
A man inspects the uniform of a figure in some East Asian country.	contradiction C C C C C	The man is sleeping
An older and younger man smiling.	neutral N N E N N	Two men are smiling and laughing at the cats playing on the floor.
A black race car starts up in front of a crowd of people.	contradiction C C C C C	A man is driving down a lonely road.

https://nlp.stanford.edu/projects/snli/

Skip the logical form and do inference from the text directly

Text	Judgments	Hypothesis
A man inspects the uniform of a figure in some East Asian country.	contradiction C C C C C	The man is sleeping
An older and younger man smiling.	neutral N N E N N	Two men are smiling and laughing at the cats playing on the floor.
A black race car starts up in front of a crowd of people.	contradiction C C C C C	A man is driving down a lonely road.
A soccer game with multiple males playing.	entailment E E E E E	Some men are playing a sport.

https://nlp.stanford.edu/projects/snli/

Skip the logical form and do inference from the text directly

Text	Judgments	Hypothesis
A man inspects the uniform of a figure in some East Asian country.	contradiction C C C C C	The man is sleeping
An older and younger man smiling.	neutral N N E N N	Two men are smiling and laughing at the cats playing on the floor.
A black race car starts up in front of a crowd of people.	contradiction C C C C C	A man is driving down a lonely road.
A soccer game with multiple males playing.	entailment E E E E E	Some men are playing a sport.
A smiling costumed woman is holding an umbrella.	neutral N N E C N	A happy woman in a fairy costume holds an umbrella.

https://nlp.stanford.edu/projects/snli/

How well are we doing?

Publication	Model	Parameters	Train (% acc)	Test (% acc)
	Feature-based models			
Bowman et al. '15	Unlexicalized features		49.4	50.4
Bowman et al. '15	+ Unigram and bigram features		99.7	78.2
	Sentence vector-based models			
Bowman et al. '15	100D LSTM encoders	220k	84.8	77.6
Bowman et al. '16	300D LSTM encoders	3.0m	83.9	80.6

How well are we doing?

Publication	Model	Parameters	Train (% acc)	Test (% acc)
	Feature-based models			
Bowman et al. '15	Unlexicalized features		49.4	50.4
Bowman et al. '15	+ Unigram and bigram features		99.7	78.2
	Sentence vector-based models			
Bowman et al. '15	100D LSTM encoders	220k	84.8	77.6
Bowman et al. '16	300D LSTM encoders	3.0m	83.9	80.6

time passes…

How well are we doing?

Publication	Model	Parameters	Train (% acc)	Test (% acc)
	Feature-based models			
Bowman et al. '15	Unlexicalized features		49.4	50.4
Bowman et al. '15	+ Unigram and bigram features		99.7	78.2
	Sentence vector-based models			
Bowman et al. '15	100D LSTM encoders	220k	84.8	77.6
Bowman et al. '16	300D LSTM encoders	3.0m	83.9	80.6

time passes…

Zhuosheng Zhang et al. '19a	SJRC (BERT-Large +SRL)	308m	95.7	91.3
Xiaodong Liu et al. '19	MT-DNN	330m	97.2	91.6
Zhousheng Zhang et al. '19b	SemBERT	339m	94.4	


SUPER
SOUL
SUNDAY

Question Answering

Question Answering

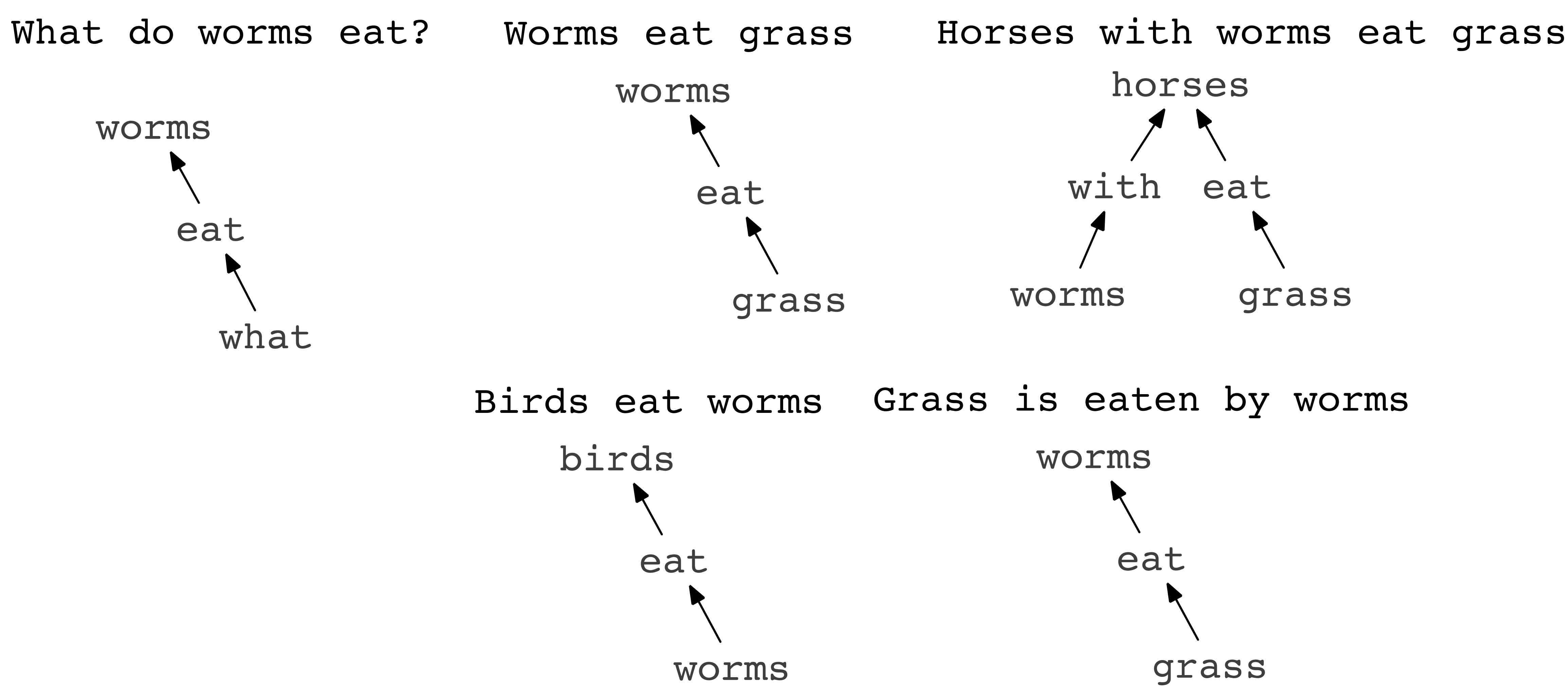
One of the oldest NLP tasks (punched card systems in 1961)

Question Answering

One of the oldest NLP tasks (punched card systems in 1961)

Simmons, Klein, McConlogue. 1964. Indexing and Dependency Logic for Answering English Questions. American Documentation 15:30, 196-204

Question: **Potential Answers:**



Question Answering: IBM's Watson

Question Answering: IBM's Watson

WILLIAM WILKINSON'S
"AN ACCOUNT OF THE
PRINCIPALITIES OF
WALLACHIA AND MOLDOVIA"
INSPIRED THIS AUTHOR'S
MOST FAMOUS NOVEL

Question Answering: IBM's Watson

WILLIAM WILKINSON'S
"AN ACCOUNT OF THE
PRINCIPALITIES OF
WALLACHIA AND MOLDOVIA"
INSPIRED THIS AUTHOR'S
MOST FAMOUS NOVEL

Bram Stoker

Question Answering: IBM's Watson

WILLIAM WILKINSON'S
"AN ACCOUNT OF THE
PRINCIPALITIES OF
WALLACHIA AND MOLDOVIA"
INSPIRED THIS AUTHOR'S
MOST FAMOUS NOVEL

Bram Stoker

- Won Jeopardy on February 16, 2011!

Apple's Siri


AT&T
22:26
65 %
“ Do I need an umbrella tomorrow in San Francisco ”
Yes, San Francisco should get rain tomorrow:
Weekly Forecast
TUES WED THU FRI SAT SUN
14° 16° 17° 17° 18° 18°
9° 7° 6° 6° 7° 8°

Wolfram Alpha

Types of Questions in Modern Systems

- Factoid questions
  - Who wrote “The Universal Declaration of Human Rights”?
  - How many calories are there in two slices of apple pie?
  - What is the average age of the onset of autism?
  - Where is Apple Computer based?

Types of Questions in Modern Systems

- Factoid questions
  - Who wrote “The Universal Declaration of Human Rights”?
  - How many calories are there in two slices of apple pie?
  - What is the average age of the onset of autism?
  - Where is Apple Computer based?

- Complex (narrative) questions:
  - In children with an acute febrile illness, what is the efficacy of acetaminophen in reducing fever?
  - What do scholars think about Jefferson’s position on dealing with pirates?

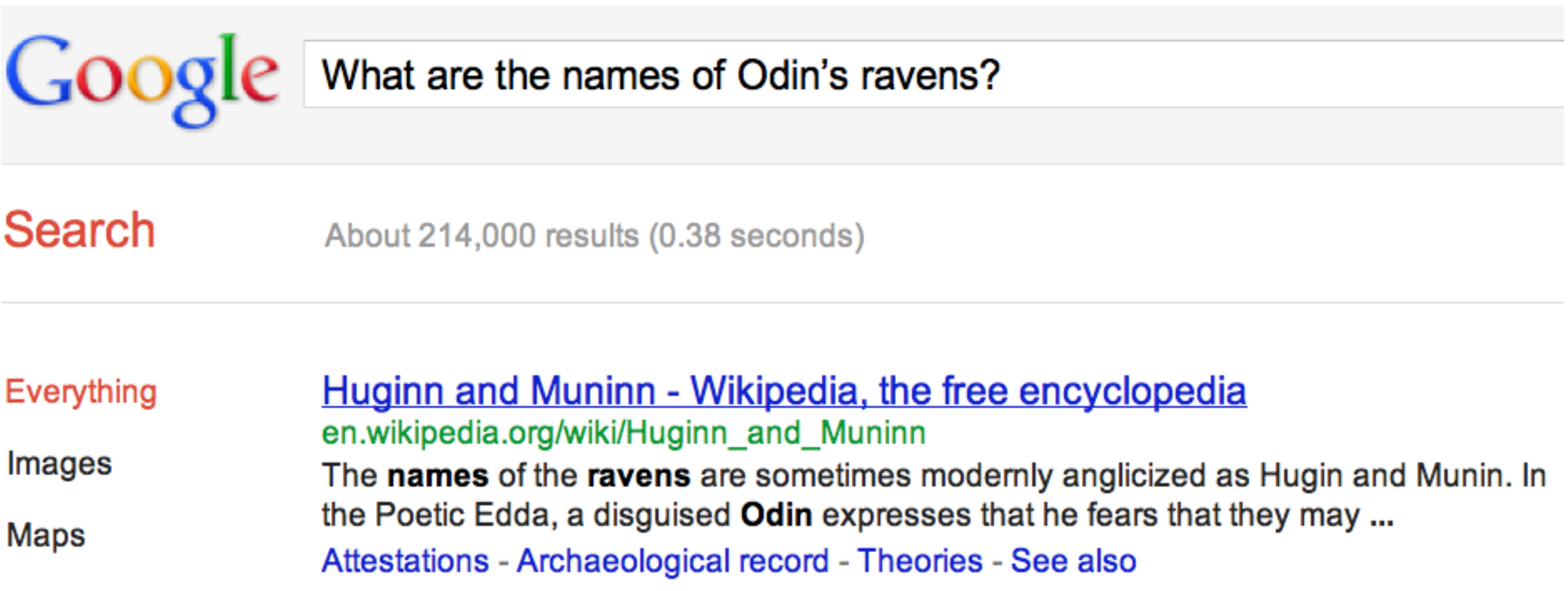
Commercial systems: mainly factoid questions

Where is the Louvre Museum located?	In Paris, France
What's the abbreviation for limited partnership?	L.P.
What are the names of Odin's ravens?	Huginn and Muninn
What currency is used in China?	The yuan
What kind of nuts are used in marzipan?	almonds
What instrument does Max Roach play?	drums
What is the telephone number for Stanford University?	650-723-2300

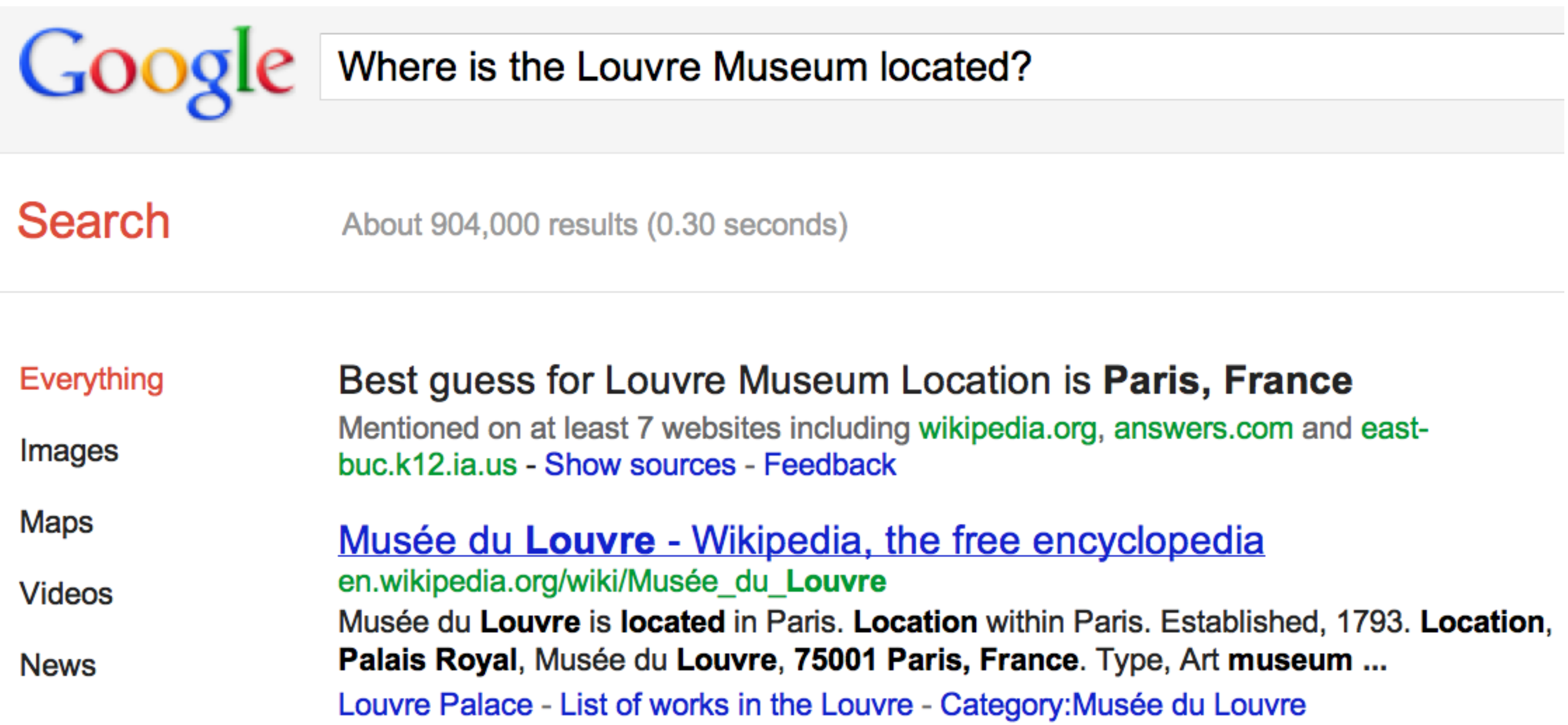
Paradigms for QA

- IR-based approaches
    - TREC; IBM Watson; Google
- Knowledge-based and Hybrid approaches
    - IBM Watson; Apple Siri; Wolfram Alpha; True Knowledge Evi

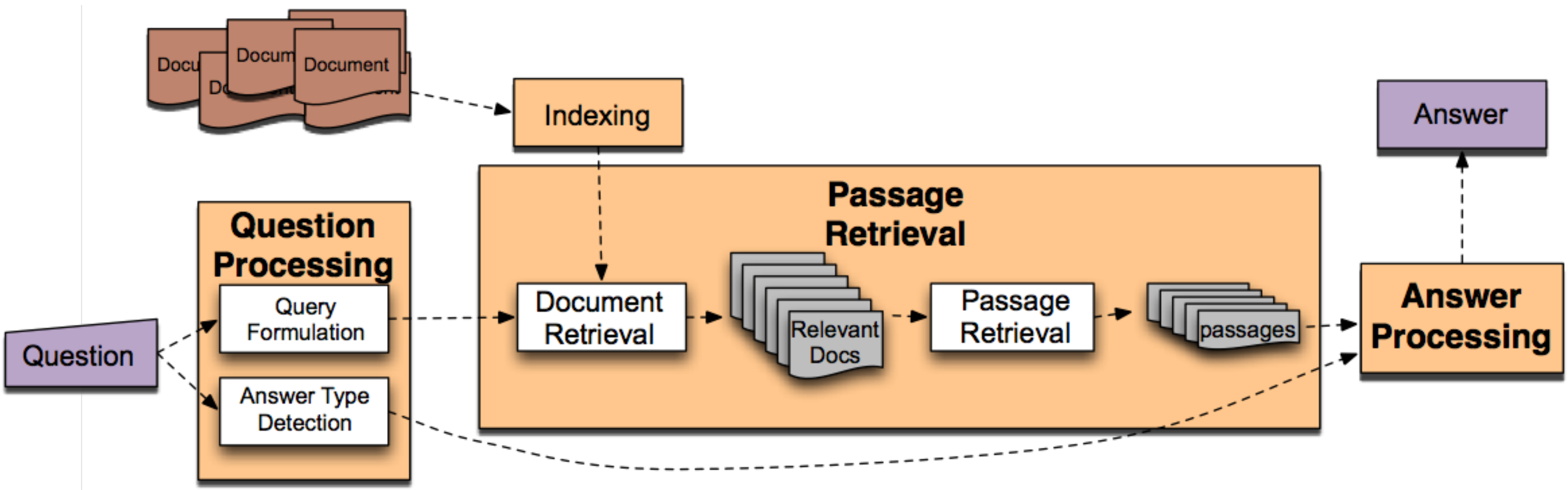
Many questions can already be answered by web search



IR-based Question Answering



IR-based Factoid QA



IR-based Factoid QA

IR-based Factoid QA

- QUESTION PROCESSING
  - Detect question type, answer type, focus, relations
  - Formulate queries to send to a search engine

IR-based Factoid QA

- QUESTION PROCESSING
  - Detect question type, answer type, focus, relations
  - Formulate queries to send to a search engine
- PASSAGE RETRIEVAL
  - Retrieve ranked documents
  - Break into suitable passages and rerank

IR-based Factoid QA

- QUESTION PROCESSING
  - Detect question type, answer type, focus, relations
  - Formulate queries to send to a search engine
- PASSAGE RETRIEVAL
  - Retrieve ranked documents
  - Break into suitable passages and rerank
- ANSWER PROCESSING
  - Extract candidate answers
  - Rank candidates
    - using evidence from the text and external sources

Knowledge-based approaches (Siri)

Knowledge-based approaches (Siri)

- Build a semantic representation of the query
  - Times, dates, locations, entities, numeric quantities

Knowledge-based approaches (Siri)

- Build a semantic representation of the query
  - Times, dates, locations, entities, numeric quantities

- Map from this semantics to query structured data or resources
  - Geospatial databases
  - Ontologies (Wikipedia infoboxes, dbPedia, WordNet, Yago)
  - Restaurant review sources and reservation services
  - Scientific databases

Hybrid approaches (IBM Watson)

Hybrid approaches (IBM Watson)

- Build a shallow semantic representation of the query

Hybrid approaches (IBM Watson)

- Build a shallow semantic representation of the query
- Generate answer candidates using IR methods
    - Augmented with ontologies and semi-structured data

Hybrid approaches (IBM Watson)

- Build a shallow semantic representation of the query
- Generate answer candidates using IR methods
  - Augmented with ontologies and semi-structured data
- Score each candidate using richer knowledge sources
  - Geospatial databases
  - Temporal reasoning
  - Taxonomical classification

Question Answering

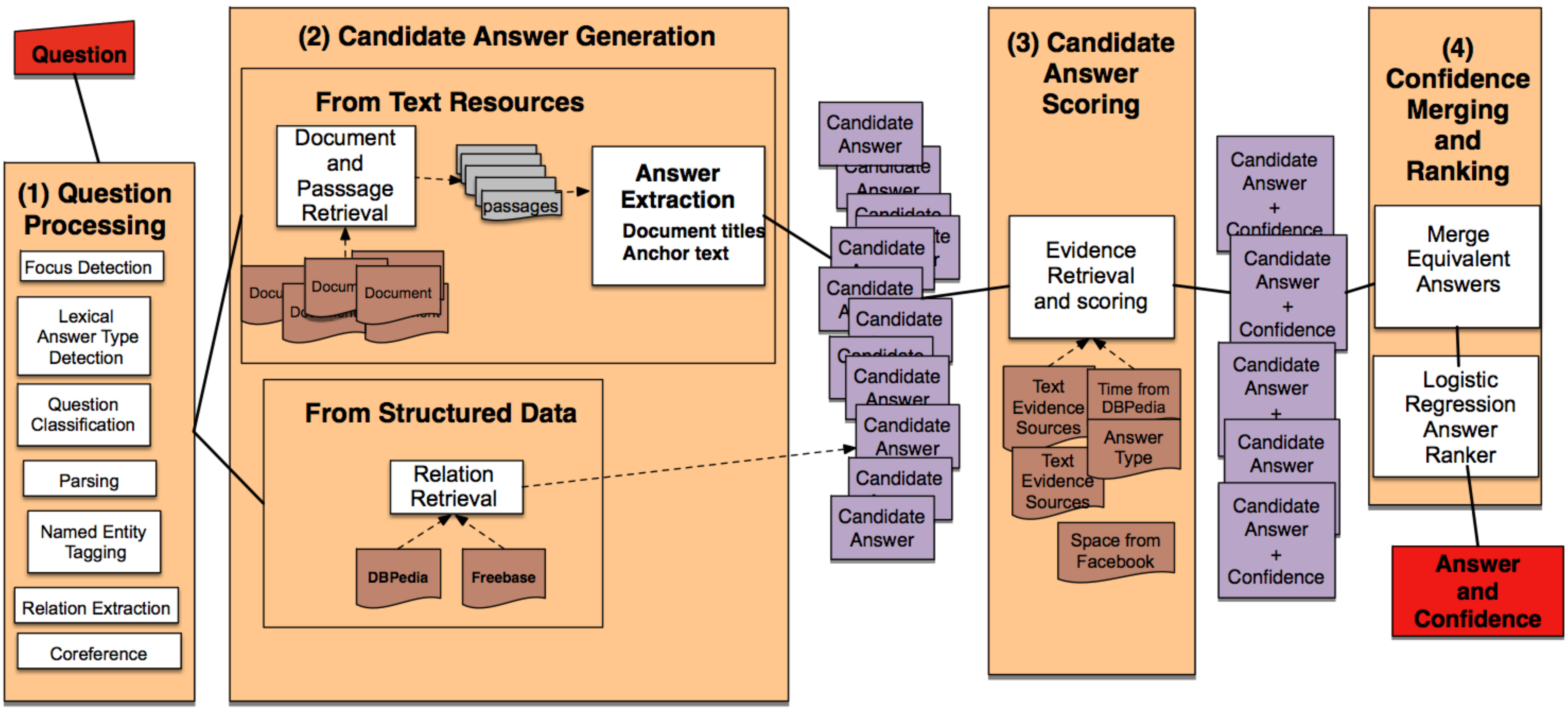
Question Answering in Watson (Deep QA)

Question Answering: IBM's Watson

WILLIAM WILKINSON'S
"AN ACCOUNT OF THE
PRINCIPALITIES OF
WALLACHIA AND MOLDOVIA"
INSPIRED THIS AUTHOR'S
MOST FAMOUS NOVEL

Bram Stoker

The Architecture of Watson



Stage 1: Question Processing

- Parsing
- Named Entity Tagging
- Relation Extraction
- Focus
- Answer Type
- Question Classification

Poets and Poetry: **He** was a bank clerk in the Yukon before he published “Songs of a Sourdough” in 1907.

GEO

Poets and Poetry: **He** was a bank clerk in the Yukon before he published "Songs of a Sourdough" in 1907.

COMPOSITION YEAR

PERSON

GEO

YEAR

GEO

Poets and Poetry: **He** was a bank clerk in the Yukon before he published "Songs of a Sourdough" in 1907.

COMPOSITION YEAR

PERSON

GEO

YEAR

Named Entity and Parse ***Focus*** ***Answer Type***

GEO

Poets and Poetry: **He** was a bank clerk in the Yukon before he published "Songs of a Sourdough" in 1907.

COMPOSITION YEAR

PERSON

Named Entity and Parse ***Focus*** ***Answer Type*** ***Relation Extraction***

GEO

Poets and Poetry: **He** was a bank clerk in the Yukon before he published "Songs of a Sourdough" in 1907.

COMPOSITION YEAR

PERSON

THEATRE: A new play based on **this Sir Arthur Conan Doyle canine classic** opened on the London stage in 2007.

GEO

YEAR

authorof(focus,"Songs of a sourdough")
publish (e1, he, "Songs of a sourdough")
in (e2, e1, 1907)
temporallink(publish(...), 1907)

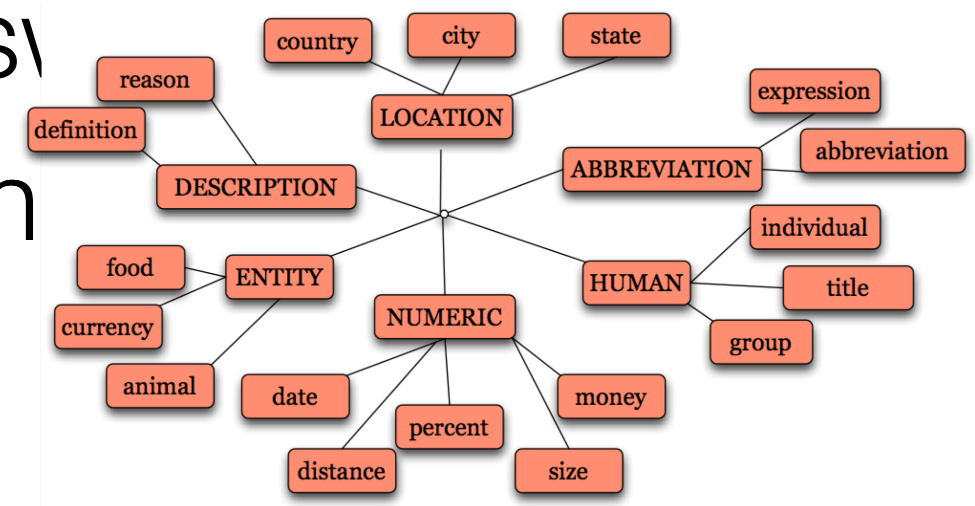
Focus extraction

- **Focus**: the part of the question that co-refers with the answer
- Replace it with answer to find a supporting passage.
- Extracted by hand-written rules
  - "Extract any noun phrase with determiner this”
  - “Extracting pronouns *she, he, hers, him,* “

Lexical Answer Type

- The semantic class of the ansv
- But for Jeopardy the TREC an

  type taxonomy is insufficient



- DeepQA team investigated 20,000 questions
- 100 named entities only covered <50% of the questions!
- Instead: Extract lots of words: 5,000 for those 20,000 questions

Lexical Answer Type

- Answer types extracted by hand-written rules
  - Syntactic headword of the focus.
  - Words that are coreferent with the focus
  - Jeopardy! category, if refers to compatible entity.

Poets and Poetry: ***He*** was a bank ***clerk*** in the Yukon before he published "Songs of a Sourdough" in 1907.

Relation Extraction in DeepQA

- For the most frequent 30 relations:
  - Hand-written regular expressions
  - AuthorOf:
    - Many patterns such as one to deal with:
    - *a Mary Shelley tale, the Saroyan novel, Twain's travel books, a1984 Tom Clancy thriller*
    - [Author] [Prose]
- For the rest: distant supervision

Stage 2: Candidate Answer Generation

Extracting candidate answers from triple stores

- If we extracted a relation from the question

  … he published “Songs of a sourdough”

```
(author-of ?x “Songs of a sourdough”)
```

Extracting candidate answers from triple stores

- If we extracted a relation from the question

  … he published “Songs of a sourdough”

```
(author-of ?x “Songs of a sourdough”)
```

- We just query a triple store
  - Wikipedia infoboxes, DBpedia, FreeBase, etc.
  - born-in(“Emma Goldman”, “June 27 1869”)
  - author-of(“Cao Xue Qin”, “Dream of the Red Chamber”)
  - author-of(“Songs of a sourdough”, “Robert Service”)

Extracting candidate answers from triple stores

- If we extracted a relation from the question

  … he published "Songs of a sourdough"

  `(author-of ?x "Songs of a sourdough")`

- We just query a triple store
  - Wikipedia infoboxes, DBpedia, FreeBase, etc.
  - born-in("Emma Goldman", "June 27 1869")
  - author-of("Cao Xue Qin", "Dream of the Red Chamber")
  - author-of("Songs of a sourdough", "Robert Service")

Extracting candidate answers from triple stores

- If we extracted a relation from the question

  … he published "Songs of a sourdough"

  `(author-of ?x "Songs of a sourdough")`

- We just query a triple store
  - Wikipedia infoboxes, DBpedia, FreeBase, etc.
  - born-in("Emma Goldman", "June 27 1869")
  - author-of("Cao Xue Qin", "Dream of the Red Chamber")
  - author-of("Songs of a sourdough", "Robert Service")

Extracting candidate answers from text: get documents/passages

1. Do standard IR-based QA to get documents

Robert Redford and Paul Newman starred in this depression-era grifter flick.

(2.0 Robert Redford) (2.0 Paul Newman) star depression era grifter (1.5 flick)

Extracting answers from documents/passages

- Useful fact: Jeopardy! answers are mostly the title of a Wikipedia document
  - If the document is a Wikipedia article, just take the title
  - If not, extract all noun phrases in the passage that are Wikipedia document titles
  - Or extract all anchor text `<a>The Sting</a>`

Stage 3: Candidate Answer Scoring

- Use lots of sources of evidence to score an answer
  - more than 50 scorers
- **Lexical answer type** is a big one
  - Different in DeepQA than in pure IR factoid QA
  - In pure IR factoid QA, answer type is used to strictly filter answers
  - In DeepQA, answer type is just one of many pieces of evidence

Lexical Answer Type (LAT) for Scoring Candidates

- Given:
  - candidate answer & lexical answer type

- Return a score: can answer can be a subclass of this answer type?

- Candidate: “***difficulty swallowing***” & LAT “***condition***”

1. Check DBPedia, WordNet, etc
   - *difficulty swallowing* -> Dbpedia *Dysphagia* -> WordNet *Dysphagia*
   - *condition*-> WordNet *Condition*

2. Check if “Dysphagia” IS-A “Condition” in WordNet
   - Wordnet for dysphagia

Relations for scoring

- **Q**: This hockey defenseman ended his career on June 5, 2008
- **Passage**: On June 5, 2008, Wesley announced his retirement after his 20th NHL season

- Question and passage have very few keywords in common
- But both have the Dbpedia relation `ActiveYearsEndDate()`

Temporal Reasoning for Scoring Candidates

- Relation databases
  - (and obituaries, biographical dictionaries, etc.)
- IBM Watson

”In 1594 he took a job as a tax collector in Andalusia”

Candidates:

- Thoreau is a bad answer (born in 1817)
- Cervantes is possible (was alive in 1594)

Geospatial knowledge (containment, directionality, borders)

- Beijing is a good answer for "Asian city"


www.geonames.org/search.html?q=palo+alto&country=

GeoNames Home | Postal Codes | Download / Webservice | About   login

palo alto   all countries

search   show on map   [advanced search]

459 records found for "palo alto"

	Name	Country	Feature class	Latitude	Longitude
1	Palo Alto Palo Al'to,Palo Alto,pa luo ao duo,paroaruto,Пало Алто,Пало Альто,פאלו אלטו,パロアルト,帕羅奧多	United States, California Santa Clara County	populated place population 64,403, elevation 9m	N 37° 26' 30''	W 122° 8' 34''
2	Palo Alto Township Palo Alto Township	United States, Iowa Jasper County	administrative division elevation 256m	N 41° 38' 15''	W 93° 2' 57''
3	Borough of Palo Alto	United States, Pennsylvania Schuylkill County	administrative division population 1,032, elevation 210m	N 40° 41' 21''	W 76° 10' 2''

Text-retrieval-based answer scorer

- Generate a query from the question and retrieve passages
- Replace the focus in the question with the candidate answer
- See how well it fits the passages.
- Robert Redford and Paul Newman starred in **this depression-era grifter flick**
- Robert Redford and Paul Newman starred in The Sting

Robert Redford - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Robert_Redford** ▾ Wikipedia ▾
Redford starred in Sydney Pollack's Out of Africa (1985), which was an .... by William Goldman, in which he was paired for the first time with **Paul Newman**. ... the blockbuster crime caper The **Sting** (1973), which became one of the top 20 ...

Stage 4: Answer Merging and Scoring

- Now we have a list candidate answers each with a score vector
  - `J.F.K  [.5  .4  1.2  33  .35 … ]`
  - `John F. Kennedy [.2 .56 5.3 2 …]`

- Merge equivalent answers*: J.F.K.* and *John F. Kennedy*
  - Use Wikipedia dictionaries that list synonyms:
    - J*FK, John F. Kennedy, John Fitzgerald Kennedy, Senator John F. Kennedy, President Kennedy, Jack Kennedy*
  - Use stemming and other morphology

Stage 4: Answer Scoring

- Build a classifier to take answers and a score vector and assign a probability
- Train on datasets of hand-labeled correct and incorrect answers.

Text Summarization


READ ALL THE ARTICLE?
AIN'T NOBODY GOT TIME FOR THAT
imgflip.com

How do we summarize this document?

MILAN, Italy, April 18. A small airplane crashed into a government building in heart of Milan, setting the top floors on fire, Italian police reported. There were no immediate reports on casualties as rescue workers attempted to clear the area in the city's financial district. Few details of the crash were available, but news reports about it immediately set off fears that it might be a terrorist act akin to the Sept. 11 attacks in the United States. Those fears sent U.S. stocks tumbling to session lows in late morning trading.

Witnesses reported hearing a loud explosion from the 30-story office building, which houses the administrative offices of the local Lombardy region and sits next to the city's central train station. Italian state television said the crash put a hole in the 25th floor of the Pirelli building. News reports said smoke poured from the opening. Police and ambulances rushed to the building in downtown Milan. No further details were immediately available.

(Radev, 2004)

How do we summarize this document?

MILAN, Italy, April 18. **A small airplane crashed** into a government building in heart of Milan, **setting the top floors on fire**, **Italian police reported**. There were **no immediate reports on casualties** as rescue workers attempted to clear the area in the city's financial district. **Few details of the crash** were available, but news reports about it immediately set off fears that it **might be a terrorist act** akin to the Sept. 11 attacks in the United States. Those fears sent **U.S. stocks tumbling** to session lows in late morning trading.

Witnesses reported hearing a loud explosion from the 30-story office building, **which houses the administrative offices of the local** Lombardy region and sits next to the city's central train station. **Italian state television** said the crash put **a hole in the 25th floor of the Pirelli building**. News reports said smoke poured from the opening. Police and ambulances rushed to the building in downtown Milan. **No further details were immediately available**.

(Radev, 2004)

How do we summarize this document?

What happened?

MILAN, Italy, April 18. **A small airplane crashed** into a government building in heart of Milan, **setting the top floors on fire, Italian police reported**. There were **no immediate reports on casualties** as rescue workers attempted to clear the area in th

How many victims?

When, where?

district. **Few de** ilabl

about it immediately set off fears that it **might be a terrorist act** akin to the Sept. 11 attacks in the United States. Those fears sent **U.S. stocks tumbling** to session lows in late morning trading.

Says who?

Was it a terrorist act?

Witnesses reported hearing a loud explosion office building, **which houses the administrative offices of the local** Lombardy region and sits next to the city's central train station. **Italian state television** said the crash put **a hole in the 25th floor of the Pirelli building**. News reports said smoke poured from the opening. Police and ambulances rushed to the building in downtown Milan. **No further details were immediately available**.

What was the target?

1. How many people were injured?
2. How many people were killed? (age, number, gender, description)
3. Was the pilot killed?
4. Where was the plane coming from?
5. Was it an accident (technical problem, illness, terrorist act)?
6. Who was the pilot? (age, number, gender, description)
7. When did the plane crash?
8. How tall is the Pirelli building?
9. Who was on the plane with the pilot?
10. Did the plane catch fire before hitting the building?
11. What was the weather like at the time of the crash?
12. When was the building built?
13. What direction was the plane flying?
14. How many people work in the building?
15. How many people were in the building at the time of the crash?
16. How many people were taken to the hospital?
17. What kind of aircraft was used?

Abstracts of papers — time saving

An Incremental Interpreter for High-Level Programs with Sensing

Giuseppe De Giacomo
Dipartimento di Informatica e Sistemistica
Universita di Roma "La Sapienza"
Via Salaria 113, 00198 Rome, Italy
degiacomo@dis.uniroma1.it

Hector Levesque
Department of Computer Science
University of Toronto
Toronto, Canada M5S 3H5
hector@cs.toronto.edu

Abstract

Like classical planning, the execution of high-level agent programs requires a reasoner to look all the way to a final goal state before even a single action can be taken in the world. This deferral is a serious problem in practice for large programs. Furthermore, the problem is compounded in the presence of sensing actions which provide necessary information, but only after they are executed in the world. To deal with this, we propose (characterize formally in the situation calculus, and implement in Prolog) a new incremental way of interpreting such high-level programs and a new high-level language construct, which together, and without loss of generality, allow much more control to be exercised over when actions can be executed. We argue that such a scheme is the only practical way to deal with large agent programs containing both nondeterminism and sensing.

Introduction

In [4] it was argued that when it comes to providing high level control to autonomous agents or robots, the notion of *high-level program execution* offers an alternative to classical planning that may be more practical in many applications. Briefly, instead of looking for a sequence of actions $\vec{a}$ such that

$$Axioms \models Legal(do(\vec{a}, S_0)) \wedge \phi(do(\vec{a}, S_0))$$

where ϕ is the goal being planned for, we look for a sequence $\vec{a}$ such that

$$Axioms \models Do(\delta, S_0, do(\vec{a}, S_0))$$

to find a sequence with the right properties. This can involve considerable search when δ is very nondeterministic, but much less search when δ is more deterministic. The feasibility of this approach for AI purposes clearly depends on the expressive power of the programming language in question. In [4], a language called ConGolog is presented, which in addition to nondeterminism, contains facilities for sequence, iteration, conditionals, concurrency, and prioritized interrupts. In this paper, we extend the expressive power of this language by providing much finer control over the nondeterminism, and by making provisions for sensing actions. To do so in a way that will be practical even for very large programs requires introducing a different style of on-line program execution.

In the rest of this section, we discuss on-line and off-line execution informally, and show why sensing actions and nondeterminism together can be problematic. In the following section, we formally characterize program execution in the language of the situation calculus. Next, we describe an incremental interpreter in Prolog that is correct with respect to this specification. The final section contains discussion and conclusions.

Off-line and On-line execution

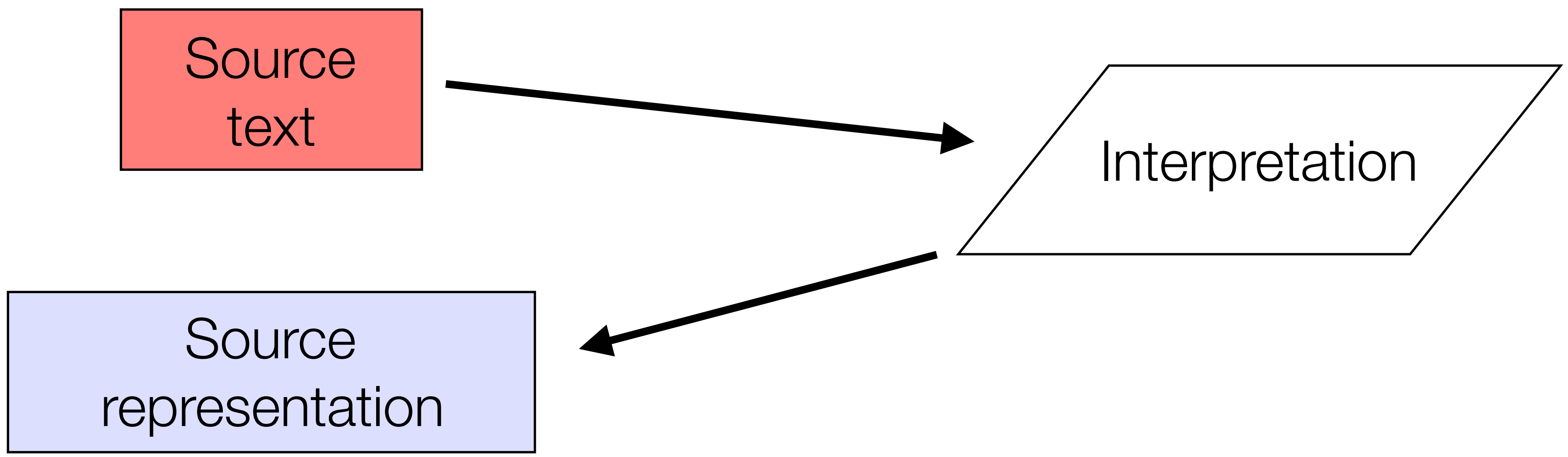
To be compatible with planning, the ConGolog interpreter presented in [4] executes in an *off-line* manner, in the sense that it must find a sequence of actions constituting an entire legal execution of a program *before* actually executing any of them in the world.[1] Consider, for example, the following program:

What exactly is text summarization?

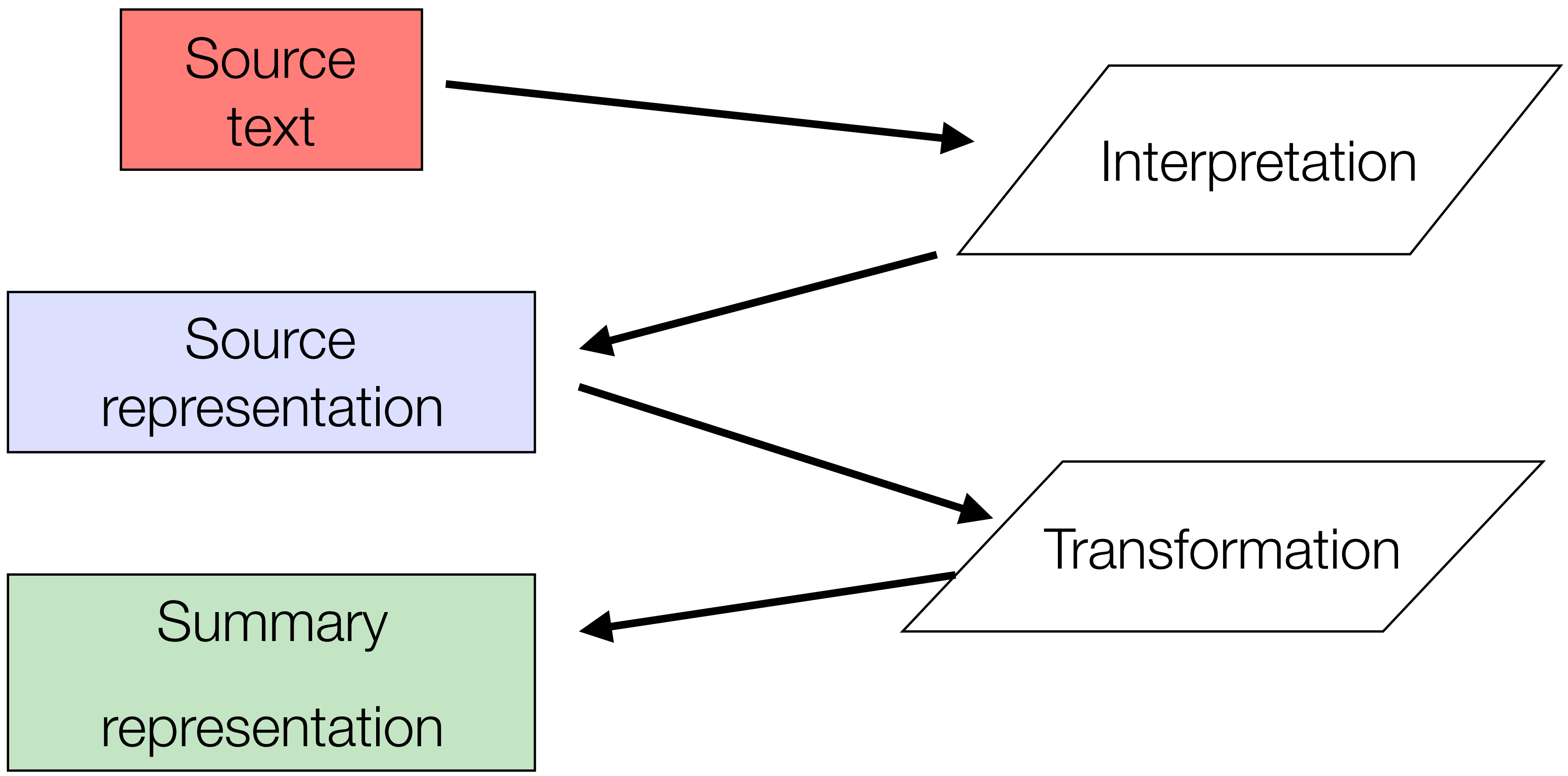
"a reductive transformation of source text to summary text through content condensation by selection and/or generalization on what is important in the source."

(Spärck Jones,1999)

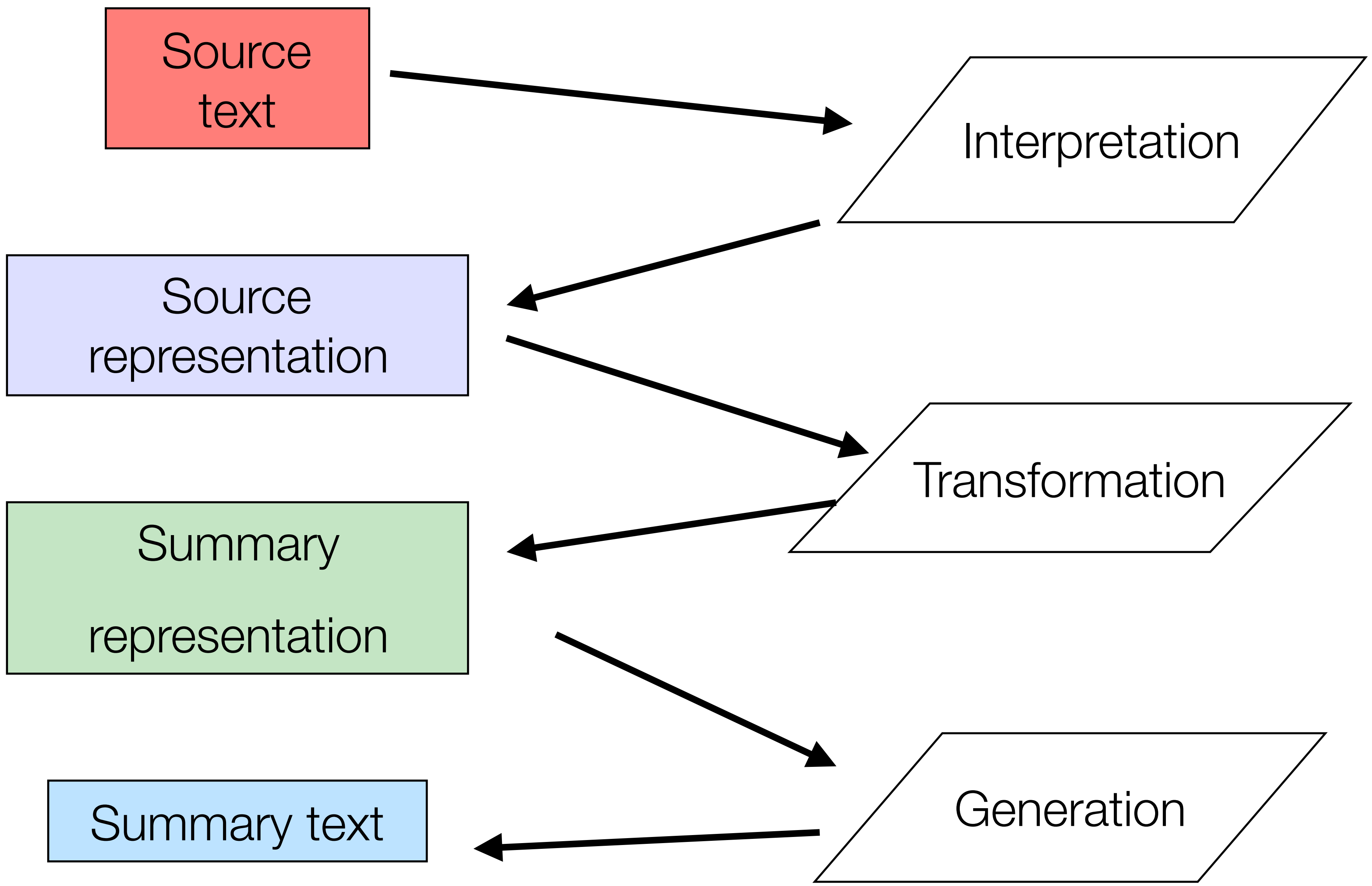
Schematic summary processing model



Schematic summary processing model



Schematic summary processing model



Summarizing factors

(Spärck Jones 2007)

Summarizing factors

- **Input**
  - subject type: domain
  - genre: newspaper articles, editorials, letters, reports...
  - form: regular text structure; free-form
  - source size: single doc; multiple docs (few; many)

(Spärck Jones 2007)

Summarizing factors

- **Input**
  - subject type: domain
  - genre: newspaper articles, editorials, letters, reports...
  - form: regular text structure; free-form
  - source size: single doc; multiple docs (few; many)
- **Purpose**
  - situation: embedded in larger system (MT, IR) or not?
  - audience: focused or general
  - usage: IR, sorting, skimming...

(Spärck Jones 2007)

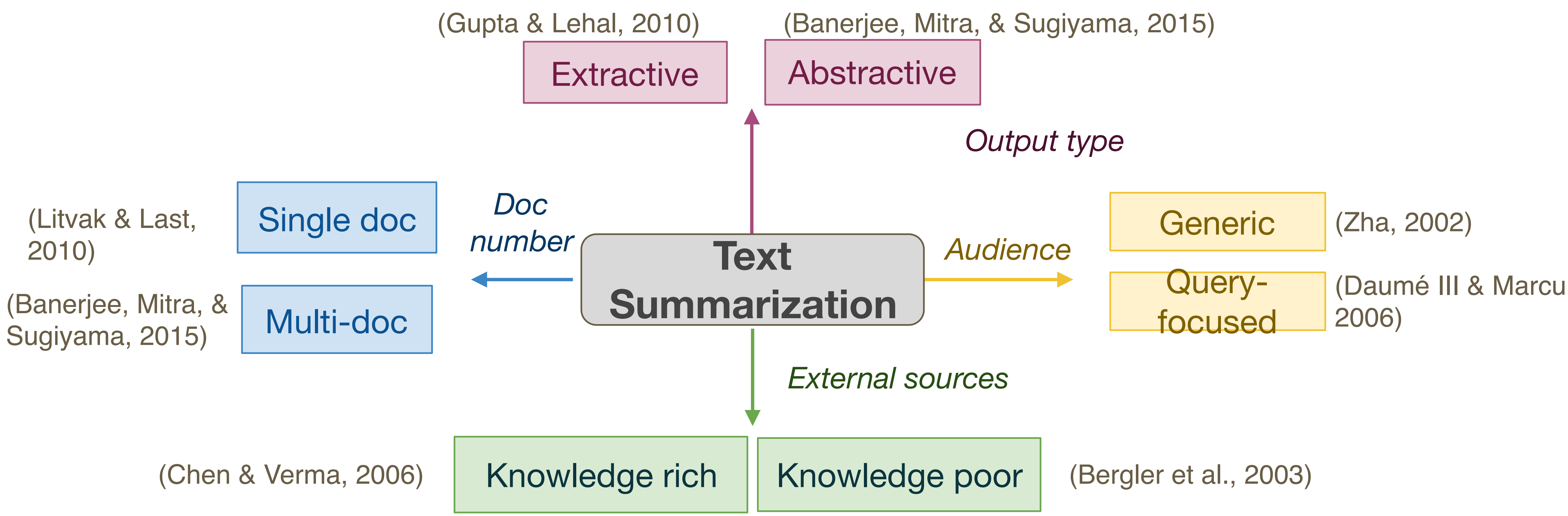
Summarizing factors

- **Input**
  - subject type: domain
  - genre: newspaper articles, editorials, letters, reports...
  - form: regular text structure; free-form
  - source size: single doc; multiple docs (few; many)
- **Purpose**
  - situation: embedded in larger system (MT, IR) or not?
  - audience: focused or general
  - usage: IR, sorting, skimming...
- **Output**
  - completeness: include all aspects, or focus on some?
  - format: paragraph, table, etc.
  - style: informative, indicative, aggregative, critical...

(Spärck Jones 2007)

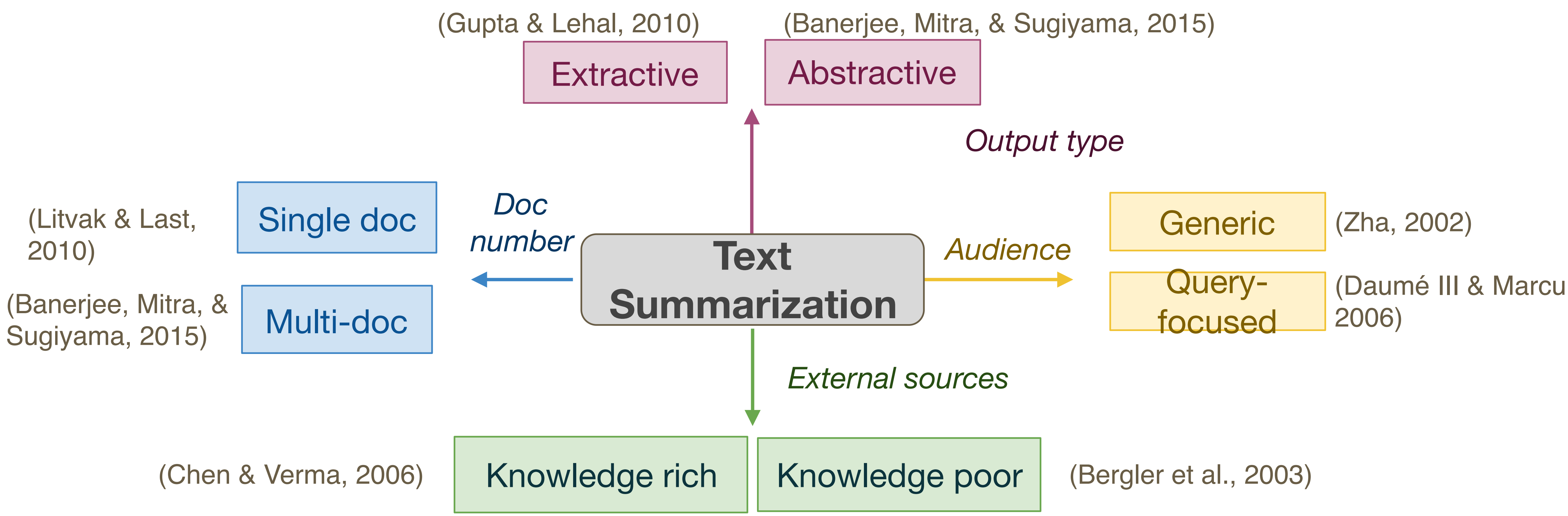
Types of Summarization

- Text Summarization Categories



Types of Summarization

- Text Summarization Categories



- Deep Learning-based Automatic Text Summarization
  - Seq2Seq model (Khatri, Singh, & Parikh, 2018)
  - Pointer-generator network (See, Liu, & Manning, 2017)

Exercise: summarize the following texts for the following readers

Exercise: summarize the following texts for the following readers

text1: Coup Attempt

Exercise: summarize the following texts for the following readers

text1: Coup Attempt

reader1: your friend, who knows nothing about South Africa.

Exercise: summarize the following texts for the following readers

text1: Coup Attempt

reader1: your friend, who knows nothing about South Africa.

reader2: someone who lives in South Africa and knows the political position.

Exercise: summarize the following texts for the following readers

text1: Coup Attempt

reader1: your friend, who knows nothing about South Africa.

reader2: someone who lives in South Africa and knows the political position.

reader3: your 4-year-old niece.

Exercise: summarize the following texts for the following readers

text1: Coup Attempt

reader1: your friend, who knows nothing about South Africa.

reader2: someone who lives in South Africa and knows the political position.

reader3: your 4-year-old niece.

reader4: the Library of Congress.

Exercise: summarize the following texts for the following readers

text1: Coup Attempt

text2: children's story

reader1: your friend, who knows nothing about South Africa.

reader2: someone who lives in South Africa and knows the political position.

reader3: your 4-year-old niece.

reader4: the Library of Congress.

90 Soldiers Arrested After Coup Attempt In Tribal Homeland
MMABATHO, South Africa (AP)

About 90 soldiers have been arrested and face possible death sentences stemming from a coup attempt in Bophuthatswana, leaders of the tribal homeland said Friday.

Rebel soldiers staged the takeover bid Wednesday, detaining homeland President Lucas Mangope and several top Cabinet officials for 15 hours before South African soldiers and police rushed to the homeland, rescuing the leaders and restoring them to power.

At least three soldiers and two civilians died in the uprising.

Bophuthatswana's Minister of Justice G. Godfrey Mothibe told a news conference that those arrested have been charged with high treason and if convicted could be sentenced to death. He said the accused were to appear in court Monday.

All those arrested in the coup attempt have been described as young troops, the most senior being a warrant officer.

During the coup rebel soldiers installed as head of state Rocky Malebane-Metsing, leader of the opposition Progressive Peoples Party.

Malebane-Metsing escaped capture and his whereabouts remained unknown, officials said. Several unsubstantiated reports said he fled to nearby Botswana.

Warrant Officer M.T.F. Phiri, described by Mangope as one of the coup leaders, was arrested Friday in Mmabatho, capital of the nominally independent homeland, officials said.

Bophuthatswana, which has a population of 1.7 million spread over seven separate land blocks, is one of 10 tribal homelands in South Africa. About half of South Africa's 26 million blacks live in the homelands, none of which are recognized internationally.

Hennie Riekert, the homeland's defense minister, said South African troops were to remain in Bophuthatswana but will not become a ``permanent presence.''

Bophuthatswana's Foreign Minister Solomon Rathebe defended South Africa's intervention.

``The fact that ... the South African government (was invited) to assist in this drama is not anything new nor peculiar to Bophuthatswana,'' Rathebe said. ``But why South Africa, one might ask? Because she is the only country with whom Bophuthatswana enjoys diplomatic relations and has formal agreements.''

Mangope described the mutual defense treaty between the homeland and South Africa as ``similar to the NATO agreement,'' referring to the Atlantic military alliance. He did not elaborate.

Asked about the causes of the coup, Mangope said, ``We granted people freedom perhaps ... to the extent of planning a thing like this.''

The uprising began around 2 a.m. Wednesday when rebel soldiers took Mangope and his top ministers from their homes to the national sports stadium.

On Wednesday evening, South African soldiers and police stormed the stadium, rescuing Mangope and his Cabinet.

South African President P.W. Botha and three of his Cabinet ministers flew to Mmabatho late Wednesday and met with Mangope, the homeland's only president since it was declared independent in 1977.

The South African government has said, without producing evidence, that the outlawed African National Congress may be linked to the coup.

The ANC, based in Lusaka, Zambia, dismissed the claims and said South Africa's actions showed that it maintains tight control over the homeland governments. The group seeks to topple the Pretoria government.

The African National Congress and other anti-government organizations consider the homelands part of an apartheid system designed to fragment the black majority and deny them political rights in South Africa.

If You Give a Mouse a Cookie

Laura Joffe Numeroff © 1985

If you give a mouse a cookie, he’s going to ask for a glass of milk.
When you give him the milk, he’ll probably ask you for a straw.
When he’s finished, he’ll ask for a napkin.
Then he’ll want to look in the mirror to make sure he doesn’t have a milk mustache.
When he looks into the mirror, he might notice his hair needs a trim.
So he’ll probably ask for a pair of nail scissors.
When he’s finished giving himself a trim, he’ll want a broom to sweep up.
He’ll start sweeping.
He might get carried away and sweep every room in the house.
He may even end up washing the floors as well.
When he’s done, he’ll probably want to take a nap.
You’ll have to fix up a little box for him with a blanket and a pillow.
He’ll crawl in, make himself comfortable, and fluff the pillow a few times.
He’ll probably ask you to read him a story.
When you read to him from one of your picture books, he'll ask to see the pictures.
When he looks at the pictures, he’ll get so excited that he’ll want to draw one of his own. He’ll ask for paper and crayons.
He’ll draw a picture. When the picture is finished, he’ll want to sign his name, with a pen.
Then he’ll want to hang his picture on your refrigerator. Which means he’ll need Scotch tape.
He’ll hang up his drawing and stand back to look at it. Looking at the refrigerator will remind him that he’s thirsty.
So…he’ll ask for a glass of milk.
And chances are that if he asks for a glass of milk, he’s going to want a cookie to go with it.

‘Genres’ of Summaries?

‘Genres’ of Summaries?

- Indicative vs. informative

  ...used for quick categorization vs. content processing.

‘Genres’ of Summaries?

- Indicative vs. informative

  ...used for quick categorization vs. content processing.

- Extract vs. abstract

  ...lists fragments of text vs. re-phrases content coherently.

‘Genres’ of Summaries?

- Indicative vs. informative

  ...used for quick categorization vs. content processing.

- Extract vs. abstract

  ...lists fragments of text vs. re-phrases content coherently.

- Generic vs. query-oriented

  ...provides author’s view vs. reflects user’s interest.

‘Genres’ of Summaries?

- Indicative vs. informative

  ...used for quick categorization vs. content processing.

- Extract vs. abstract

  ...lists fragments of text vs. re-phrases content coherently.

- Generic vs. query-oriented

  ...provides author’s view vs. reflects user’s interest.

- Background vs. just-the-news

  ...assumes reader’s prior knowledge is poor vs. up-to-date.

‘Genres’ of Summaries?

- Indicative vs. informative

  ...used for quick categorization vs. content processing.

- Extract vs. abstract

  ...lists fragments of text vs. re-phrases content coherently.

- Generic vs. query-oriented

  ...provides author’s view vs. reflects user’s interest.

- Background vs. just-the-news

  ...assumes reader’s prior knowledge is poor vs. up-to-date.

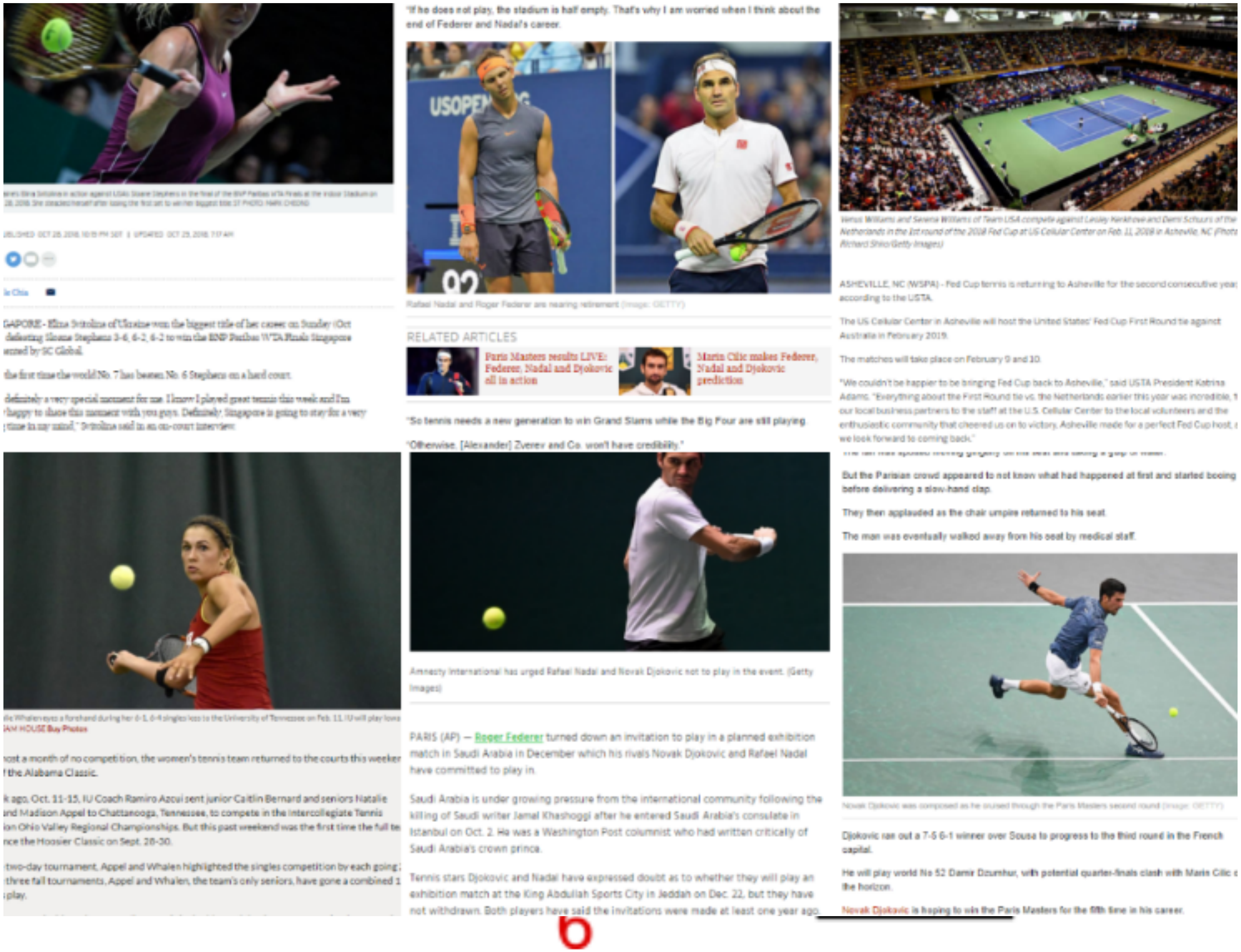
- Single-document vs. multi-document source

  ...based on one text vs. fuses together many texts.

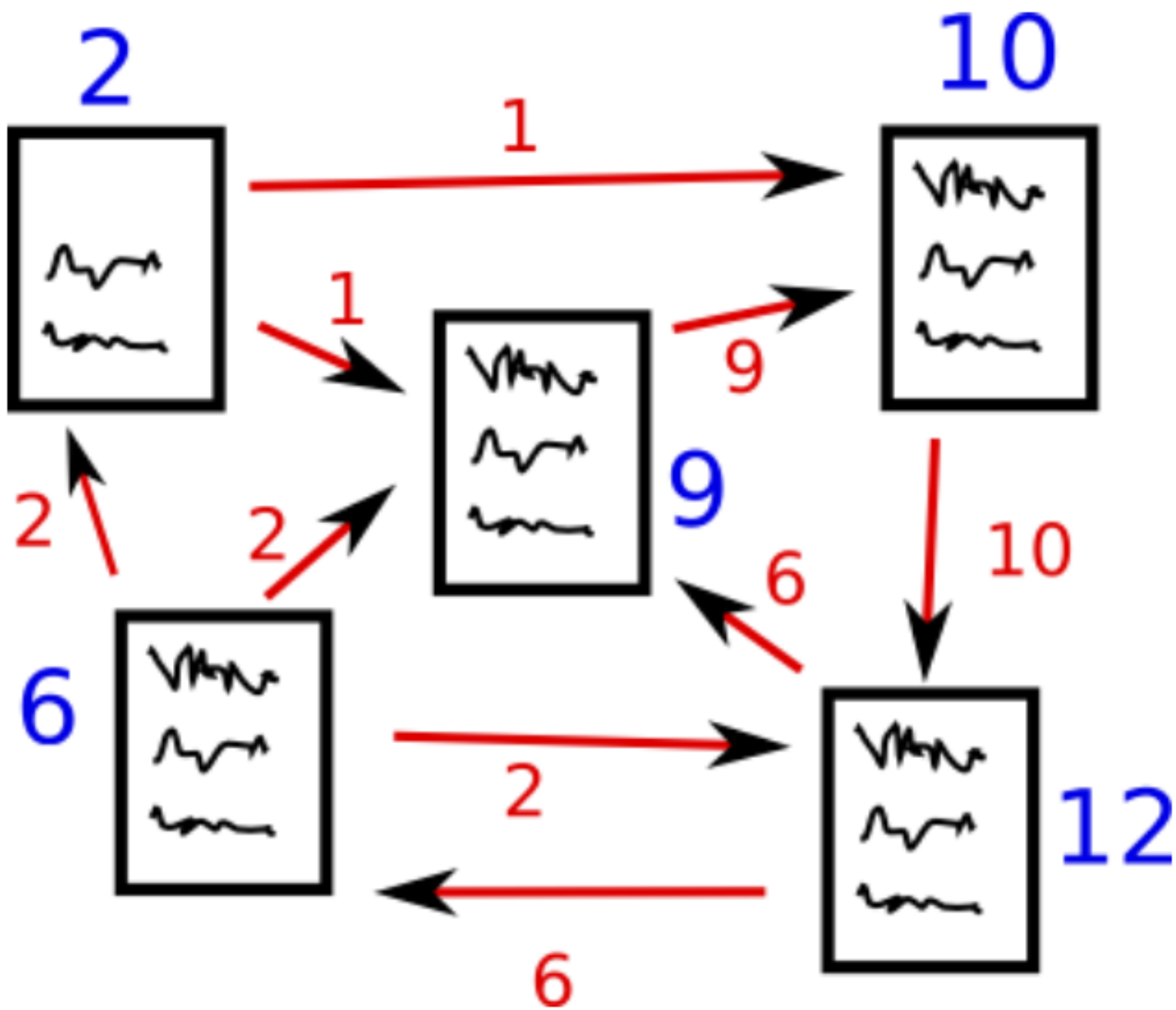
How do we do summarization?

- Old school methods
- Graph-based methods
- Neural methods

Graph based methods try to represent text as a network



Graph based methods try to represent text as a network



A single document can be a graph too!

- Each sentence is a node in the graph
- Add edges between sentences that have some similarity
- Use some threshold *t* to decide when drop edges

PageRank can tell us which documents or sentences are more important

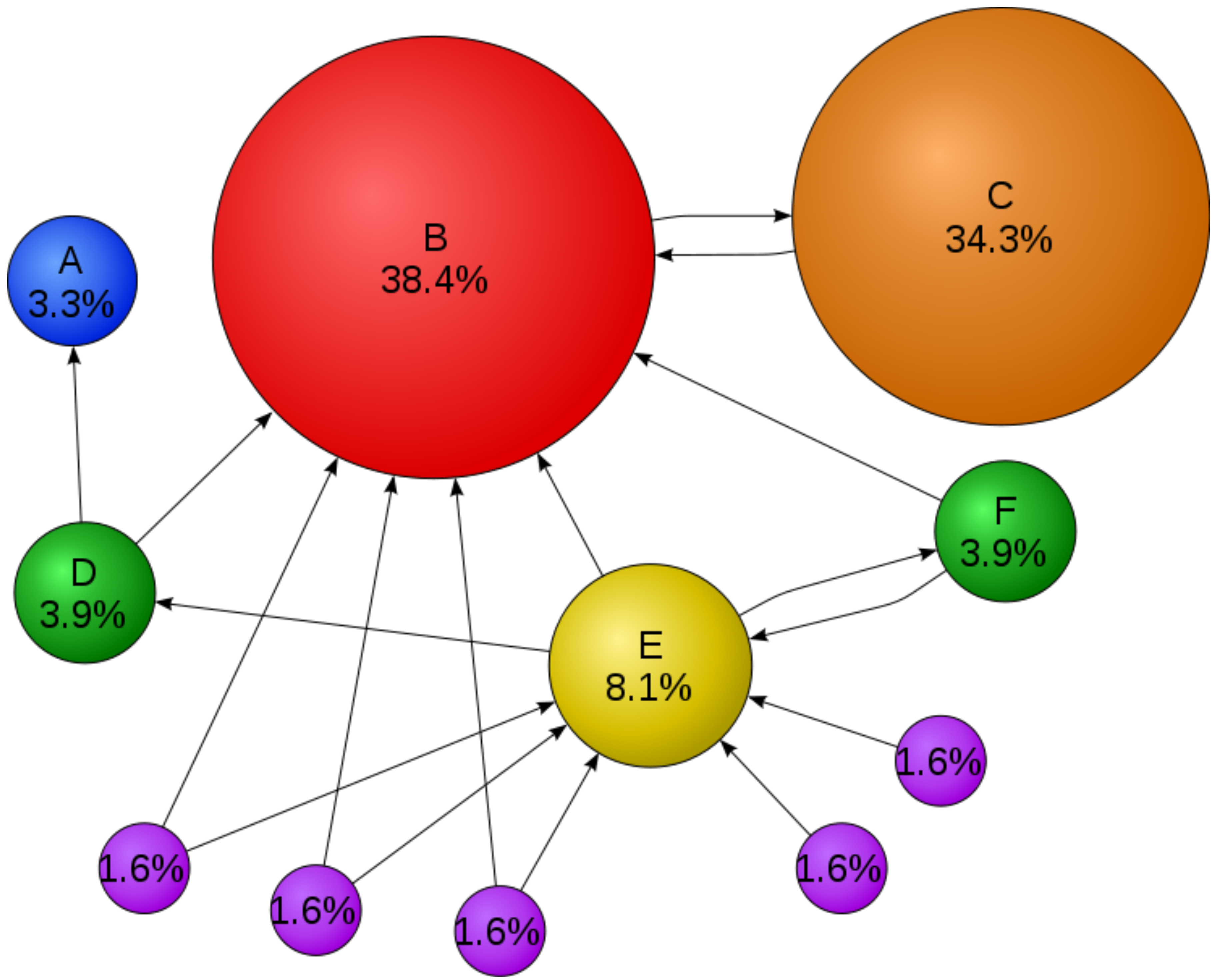


Image courtesy: Wikipedia

Text as a graph

Text as a graph

- Sentences in the text are modeled as vertices of the graph.

Text as a graph

- Sentences in the text are modeled as vertices of the graph.
- Two vertices are connected if there exists a similarity relation between them.

Text as a graph

- Sentences in the text are modeled as vertices of the graph.
- Two vertices are connected if there exists a similarity relation between them.
- Similarity formula

$$Similarity(S_i, S_j) = \frac{|\{w_k | w_k \in S_i \& w_k \in S_j\}|}{log(|S_i|) + log(|S_j|)}$$

Text as a graph

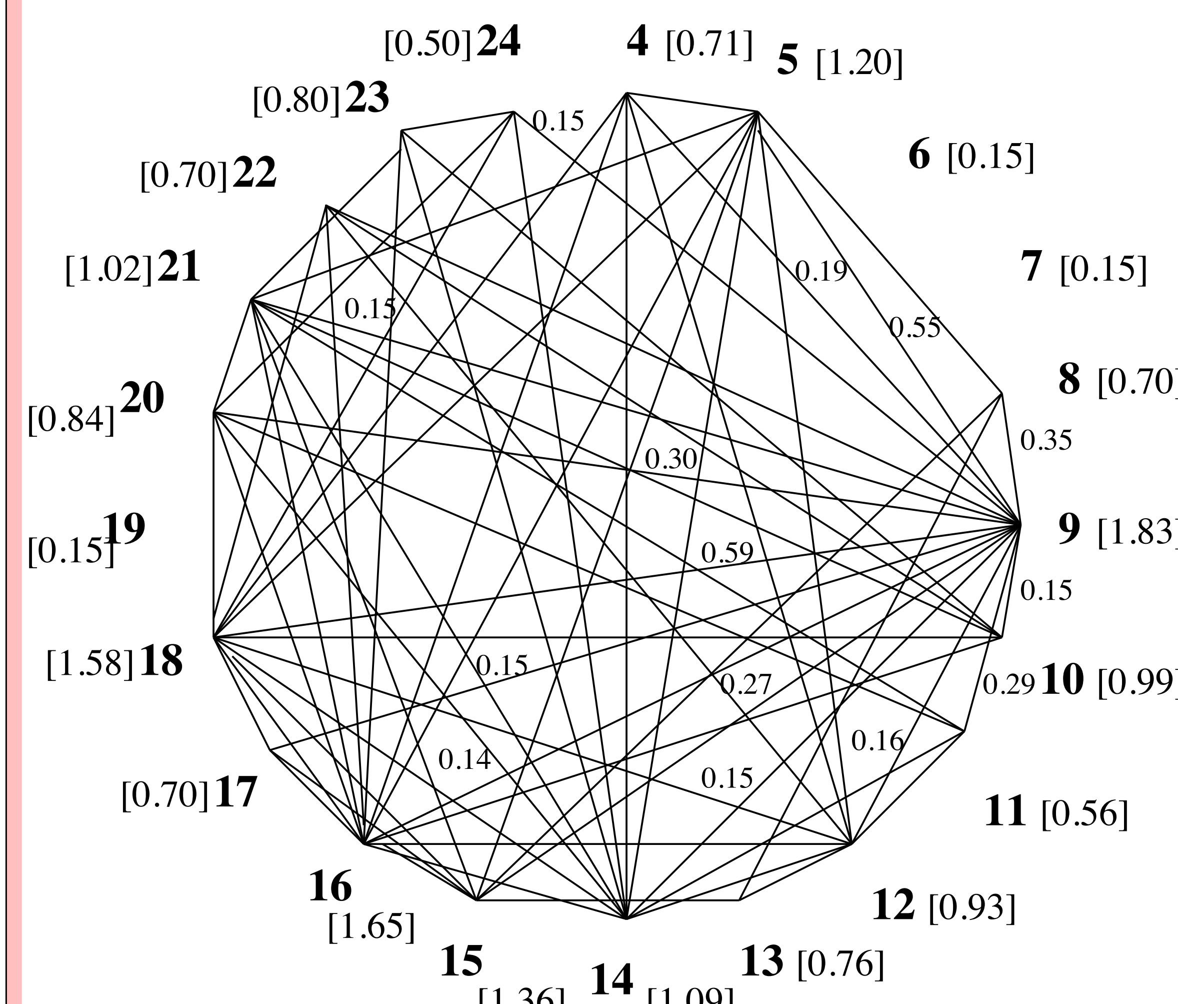
- Sentences in the text are modeled as vertices of the graph.
- Two vertices are connected if there exists a similarity relation between them.
- Similarity formula

$$Similarity(S_i, S_j) = \frac{|\{w_k | w_k \in S_i \& w_k \in S_j\}|}{log(|S_i|) + log(|S_j|)}$$

- After the ranking algorithm is run on the graph, sentences are sorted in reversed order of their score, and the top ranked sentences are selected.

TextRank example

3: BC−HurricaineGilbert, 09−11 339
4: BC−Hurricaine Gilbert, 0348
5: Hurricaine Gilbert heads toward Dominican Coast
6: By Ruddy Gonzalez
7: Associated Press Writer
8: Santo Domingo, Dominican Republic (AP)
9: Hurricaine Gilbert Swept towrd the Dominican Republic Sunday, and the Civil Defense alerted its heavily populated south coast to prepare for high winds, heavy rains, and high seas.
10: The storm was approaching from the southeast with sustained winds of 75 mph gusting to 92 mph.
11: "There is no need for alarm," Civil Defense Director Eugenio Cabral said in a television alert shortly after midnight Saturday.
12: Cabral said residents of the province of Barahona should closely follow Gilbert’s movement.
13: An estimated 100,000 people live in the province, including 70,000 in the city of Barahona, about 125 miles west of Santo Domingo.
14. Tropical storm Gilbert formed in the eastern Carribean and strenghtened into a hurricaine Saturday night.
15: The National Hurricaine Center in Miami reported its position at 2 a.m. Sunday at latitude 16.1 north, longitude 67.5 west, about 140 miles south of Ponce, Puerto Rico, and 200 miles southeast of Santo Domingo.
16: The National Weather Service in San Juan, Puerto Rico, said Gilbert was moving westard at 15 mph with a "broad area of cloudiness and heavy weather" rotating around the center of the storm.
17. The weather service issued a flash flood watch for Puerto Rico and the Virgin Islands until at least 6 p.m. Sunday.
18: Strong winds associated with the Gilbert brought coastal flooding, strong southeast winds, and up to 12 feet to Puerto Rico’s south coast.
19: There were no reports on casualties.
20: San Juan, on the north coast, had heavy rains and gusts Saturday, but they subsided during the night.
21: On Saturday, Hurricane Florence was downgraded to a tropical storm, and its remnants pushed inland from the U.S. Gulf Coast.
22: Residents returned home, happy to find little damage from 90 mph winds and sheets of rain.
23: Florence, the sixth named storm of the 1988 Atlantic storm season, was the second hurricane.
24: The first, Debby, reached minimal hurricane strength briefly before hitting the Mexican coast last month.



TextRank example

3: BC−HurricaineGilbert, 09−11 339
4: BC−Hurricaine Gilbert, 0348
5: Hurricaine Gilbert heads toward Dominican Coast
6: By Ruddy Gonzalez
7: Associated Press Writer
8: Santo Domingo, Dominican Republic (AP)
9: Hurricaine Gilbert Swept towrd the Dominican Republic Sunday, and the Civil Defense alerted its heavily populated south coast to prepare for high winds, heavy rains, and high seas.
10: The storm was approaching from the southeast with sustained winds of 75 mph gusting to 92 mph.
11: "There is no need for alarm," Civil Defense Director Eugenio Cabral said in a television alert shortly after midnight Saturday.
12: Cabral said residents of the province of Barahona should closely follow Gilbert's movement.
13: An estimated 100,000 people live in the province, including 70,000 in the city of Barahona, about 125 miles west of Santo Domingo.
14. Tropical storm Gilbert formed in the eastern Carribean and strenghtened into a hurricaine Saturday night.
15: The National Hurricaine Center in Miami reported its position at 2 a.m. Sunday at latitude 16.1 north, longitude 67.5 west, about 140 miles south of Ponce, Puerto Rico, and 200 miles southeast of Santo Domingo.
16: The National Weather Service in San Juan, Puerto Rico, said Gilbert was moving westard at 15 mph with a "broad area of cloudiness and heavy weather" rotating around the center of the storm.
17. The weather service issued a flash flood watch for Puerto Rico and the Virgin Islands until at least 6 p.m. Sunday.
18: Strong winds associated with the Gilbert brought coastal flooding, strong southeast winds, and up to 12 feet to Puerto Rico's south coast.
19: There were no reports on casualties.
20: San Juan, on the north coast, had heavy rains and gusts Saturday, but they subsided during the night.
21: On Saturday, Hurricane Florence was downgraded to a tropical storm, and its remnants pushed inland from the U.S. Gulf Coast.
22: Residents returned home, happy to find little damage from 90 mph winds and sheets of rain.
23: Florence, the sixth named storm of the 1988 Atlantic storm season, was the second hurricane.
24: The first, Debby, reached minimal hurricane strength briefly before hitting the Mexican coast last month.

- **TextRank extractive summary**
  - Hurricane Gilbert swept toward the Dominican Republic Sunday, and the Civil Defense alerted its heavily populated south coast to prepare for high winds, heavy rains and high seas. The National Hurricane Center in Miami reported its position at 2 a.m. Sunday at latitude 16.1 north, longitude 67.5 west, about 140 miles south of Ponce, Puerto Rico, and 200 miles southeast of Santo Domingo. The National Weather Service in San Juan, Puerto Rico, said Gilbert was moving westward at 15 mph with a "broad area of cloudiness and heavy weather" rotating around the center of the storm. Strong winds associated with Gilbert brought coastal flooding, strong southeast winds and up to 12 feet to Puerto Rico's south coast.
- **Manual abstract I**
  - Hurricane Gilbert is moving toward the Dominican Republic, where the residents of the south coast, especially the Barahona Province, have been alerted to prepare for heavy rains, and high wind and seas. Tropical storm Gilbert formed in the eastern Caribbean and became a hurricane on Saturday night. By 2 a.m. Sunday it was about 200 miles southeast of Santo Domingo and moving westward at 15 mph with winds of 75 mph. Flooding is expected in Puerto Rico and in the Virgin Islands. The second hurricane of the season, Florence, is now over the southern United States and down– graded to a tropical storm.
- **Manual abstract II**
  - Tropical storm Gilbert in the eastern Caribbean strengthened into a hurricane Saturday night. The National Hurricane Center in Miami reported its position at 2 a.m. Sunday to be about 140 miles south of Puerto Rico and 200 miles southeast of Santo Domingo. It is moving westward at 15 mph with a broad area of cloudiness and heavy weather with sustained winds of 75 mph gusting to 92 mph. The Dominican Republic's Civil Defense alerted that country's heavily populated south coast and the National Weather Service in San Juan, Puerto Rico issued a flood watch for Puerto Rico and the Virgin Islands until at least 6 p.m. Sunday.

Why TextRank works?

Why TextRank works?

- Through the graphs, TextRank identifies connections between various entities, and implements the concept of recommendation.

Why TextRank works?

- Through the graphs, TextRank identifies connections between various entities, and implements the concept of recommendation.
- A text unit recommends other related text units, and the strength of the recommendation is recursively computed based on the importance of the units making the recommendation.

Why TextRank works?

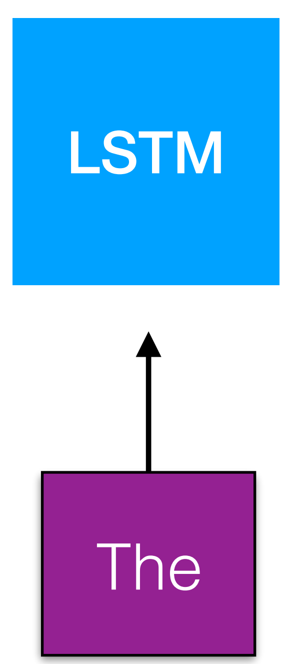
- Through the graphs, TextRank identifies connections between various entities, and implements the concept of recommendation.
- A text unit recommends other related text units, and the strength of the recommendation is recursively computed based on the importance of the units making the recommendation.
- The sentences that are highly recommended by other sentences in the text are likely to be more informative

How do we do summarization?

- Old school methods
- Graph-based methods
- Neural methods

Summarization can be thought of as a sequence to sequence task

Reminder:
a LSTM is a type of recurrent neural network

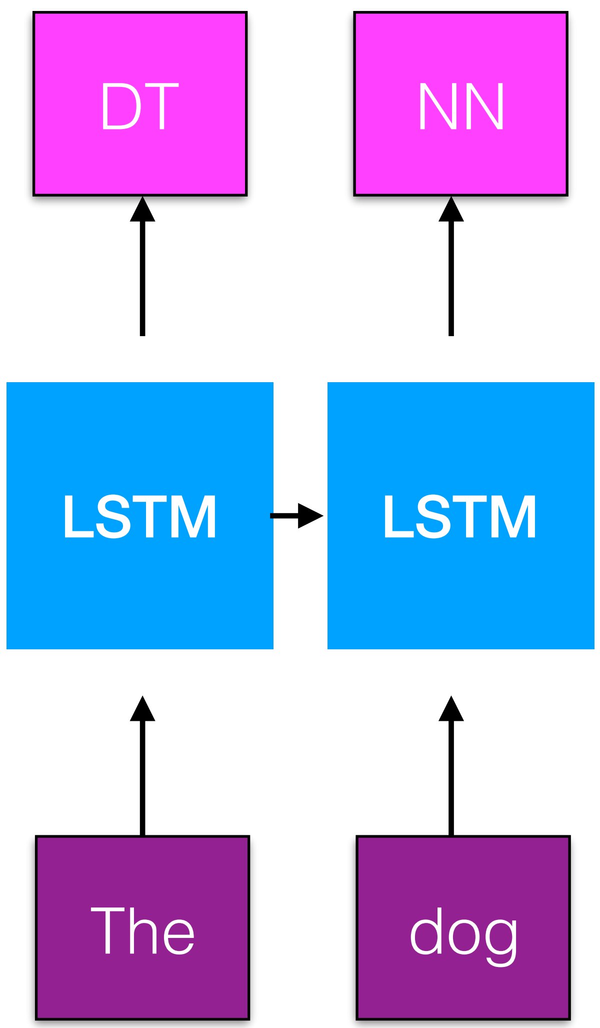


Summarization can be thought of as a sequence to sequence task

Reminder:
a LSTM is a type of recurrent neural network

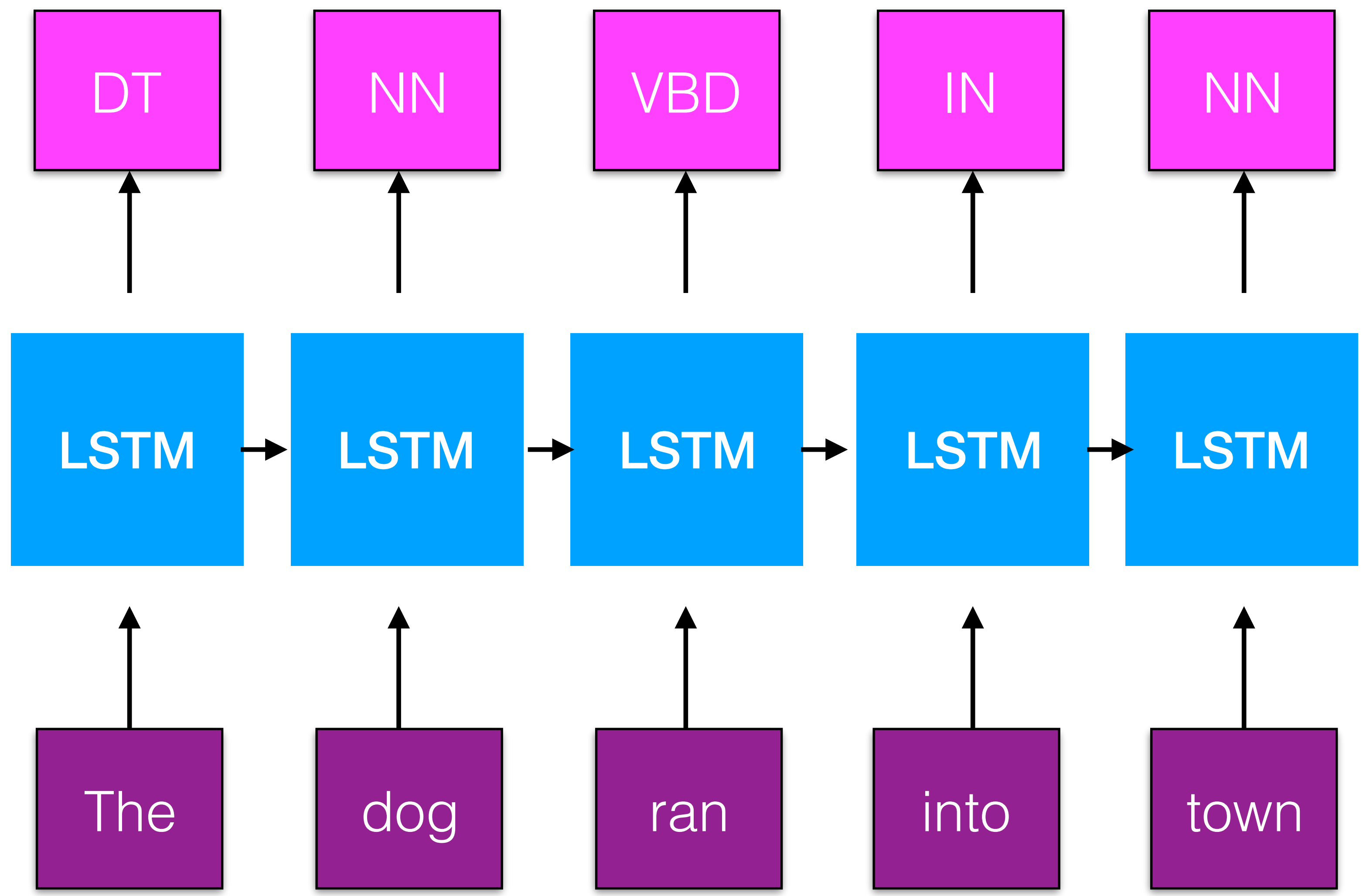
Summarization can be thought of as a sequence to sequence task

Reminder:
a LSTM is a type of recurrent neural network



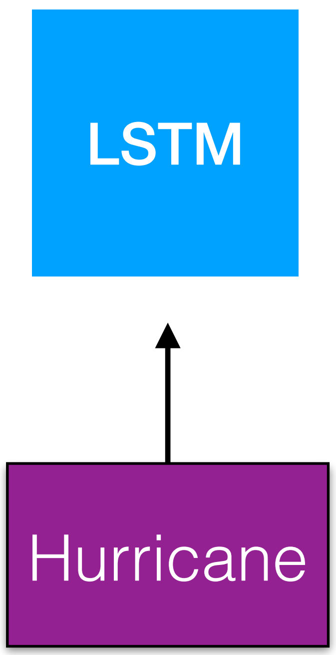
Summarization can be thought of as a sequence to sequence task

Reminder:
a LSTM is a type of recurrent neural network



Summarization can be thought of as a sequence to sequence task

Reminder:
a LSTM is a type of recurrent neural network

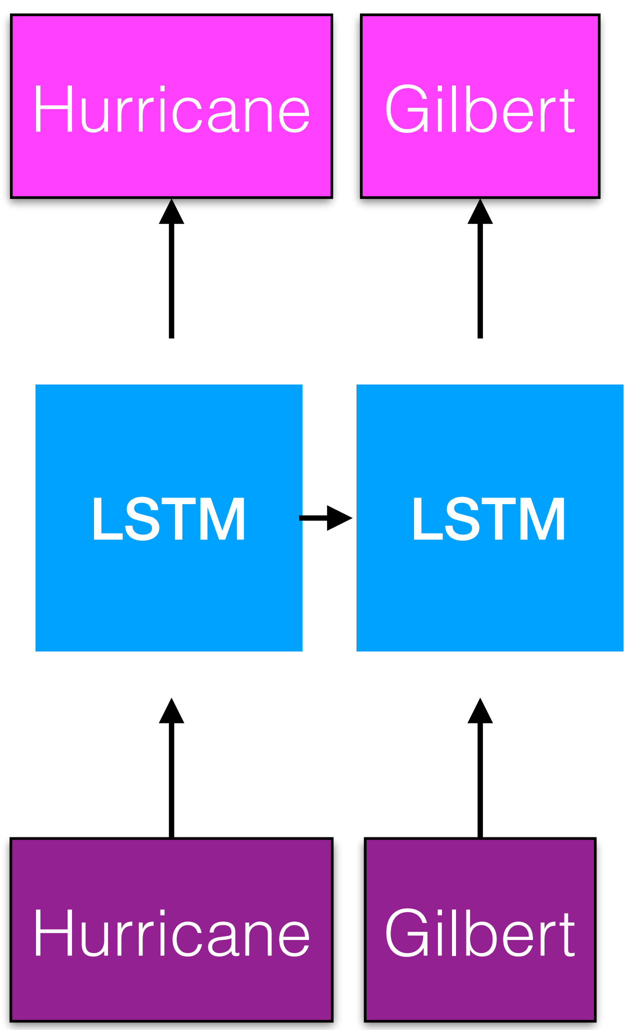


Summarization can be thought of as a sequence to sequence task

Reminder:
a LSTM is a type of recurrent neural network

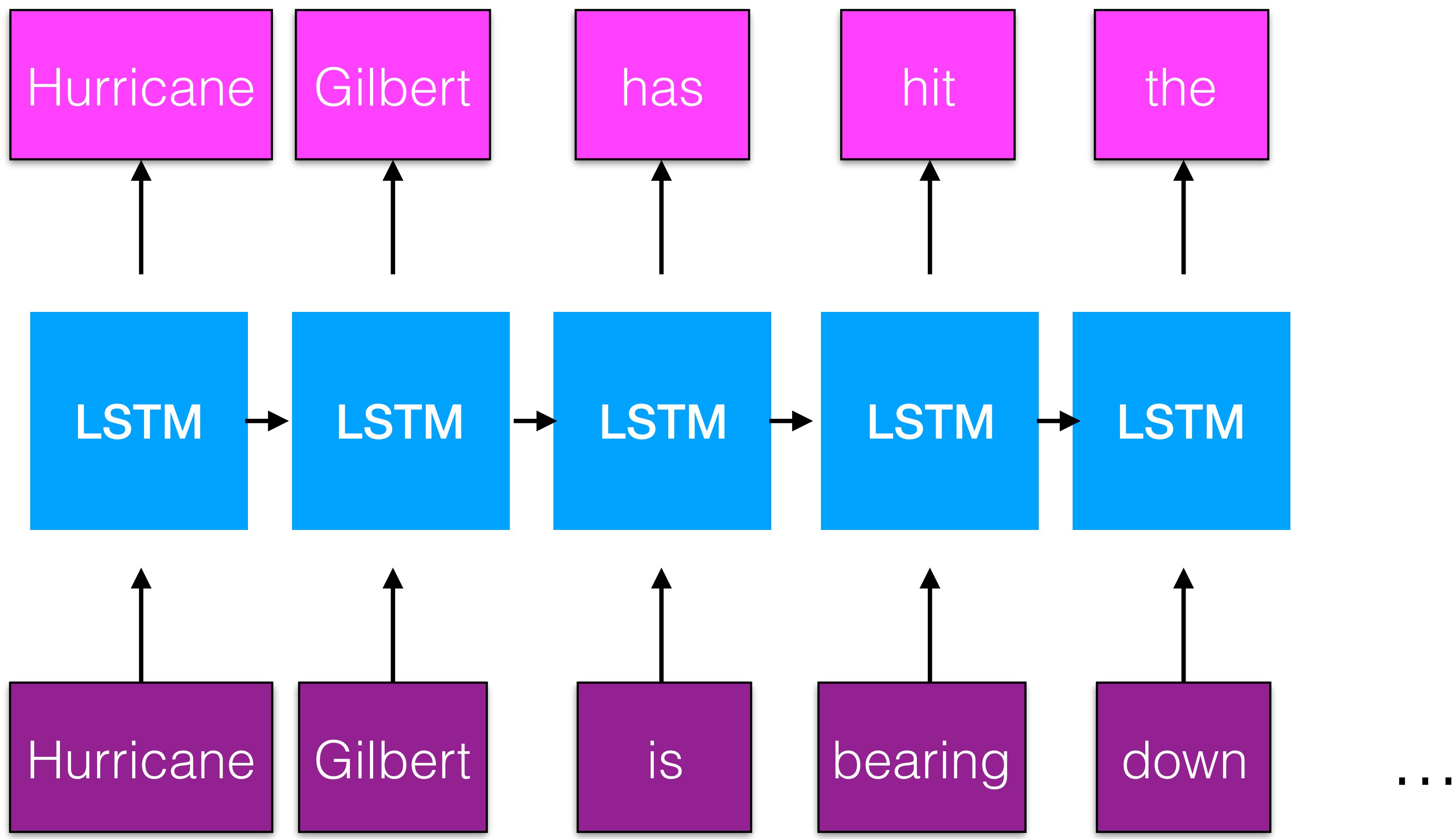
Summarization can be thought of as a sequence to sequence task

Reminder:
a LSTM is a type of recurrent neural network



Summarization can be thought of as a sequence to sequence task

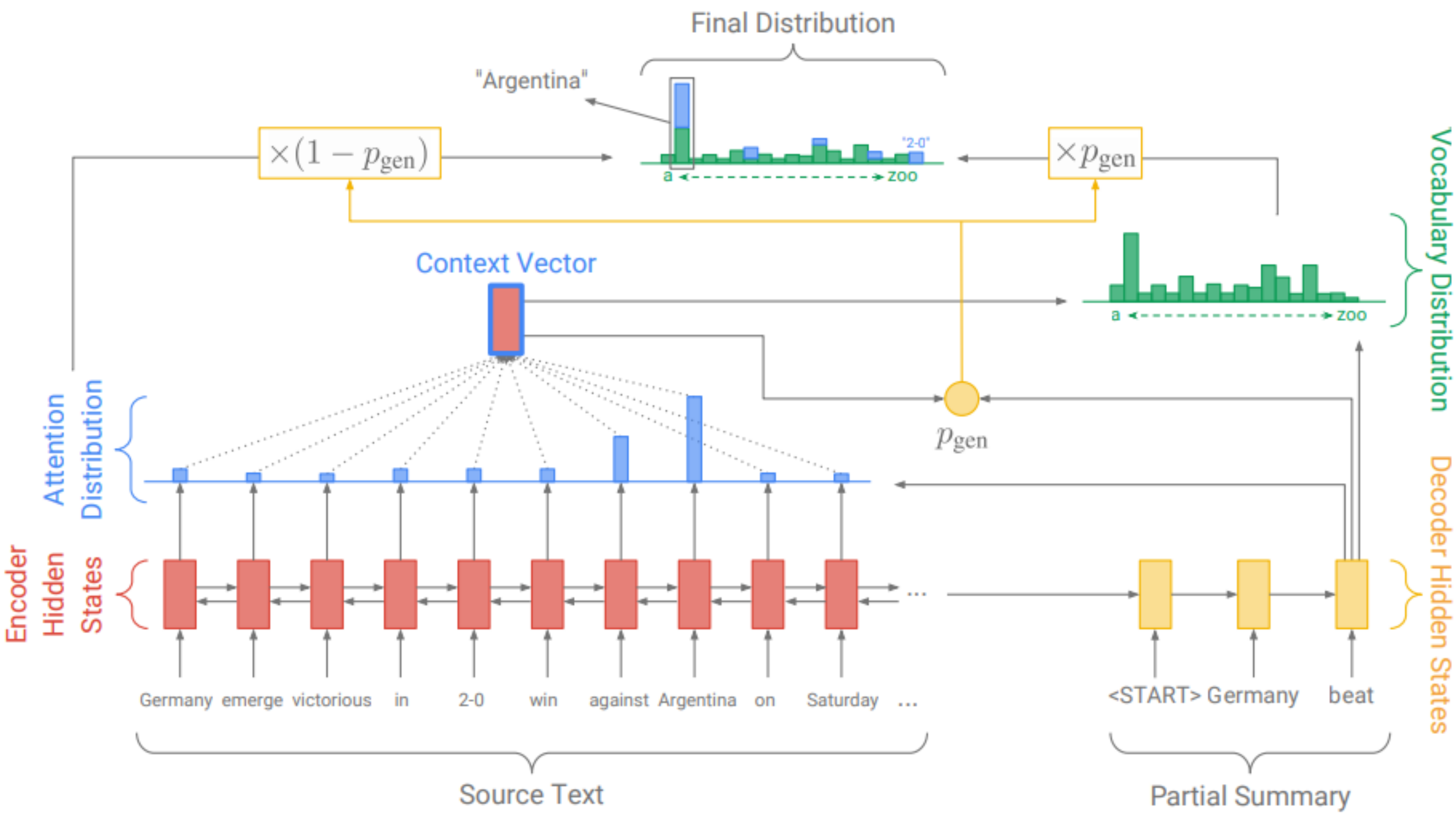
Reminder:
a LSTM is a type of recurrent neural network



Seq2Seq for Summarization

- Advantages
  - Great of abstractive summarization
- Disadvantages
  - Limited vocabulary size
  - Need lots of training data

One solution for out of vocab words: Pointer-Generator Network



Pointer-Generator Network

- How to deal with OOV words?
- Learn $p_{gen} = \sigma(w_c c_i + w_s s_i + w_x x_i + b)$
- p_{gen} is used as a soft switch to choose between generating a word from the vocabulary by sampling from P_{vocab}, or copying a word from the input sequence by sampling from the attention distribution a_i.
- Sample from the extend vocabulary distribution:

$$P(w) = p_{gen} P_{vocab}(w) + (1 - p_{gen}) \sum_{j:w_j=w} a_j^i$$

Vinyals et al., 2015

Coverage Mechanism

- Repetition is a common problem for sequence-to-sequence models
- Maintain coverage vector $cov_i = \sum_{j=0}^{i-1} a_j$
  - The (unnormalized) distribution over the source document words
- In order to ensure the attention mechanism is informed of its previous decisions Add this to the alignment model:
  $e_{ij} = v^T \tanh\left(Ws_{i-1} + Uh_j + Vcov_i + b_{attn}\right)$
- Loss function
  - $Loss_i = -\log P\left(w_i^*\right) + \lambda \sum_j \min(a_{ij}, cov_{ij})$

Tu et al., 2016

Results

- ROUGE scores of the trained models on the CNN-Daily Mail dataset, computed on the test set.

	ROUGE			METEOR	
	1	2	L	exact match	+ stem/syn/para
abstractive model (Nallapati et al., 2016)*	35.46	13.30	32.65	-	-
seq-to-seq + attn baseline (150k vocab)	30.49	11.17	28.08	11.65	12.86
seq-to-seq + attn baseline (50k vocab)	31.33	11.81	28.83	12.03	13.20
pointer-generator	36.44	15.66	33.42	15.35	16.65
pointer-generator + coverage	**39.53**	**17.28**	**36.38**	17.32	18.72
lead-3 baseline (ours)	40.34	17.70	36.57	20.48	22.21
lead-3 baseline (Nallapati et al., 2017)*	39.2	15.7	35.5	-	-
extractive model (Nallapati et al., 2017)*	39.6	16.2	35.3	-	-

How can You Evaluate a Summary?

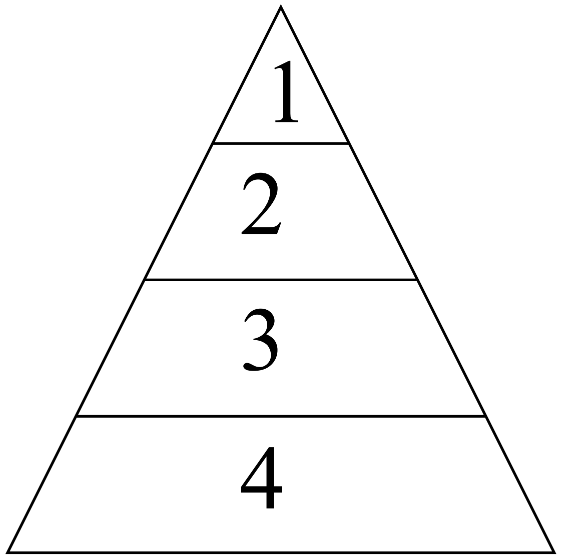
- **When you generate a summary…**

  1. the gold standard summaries (human),
  2. choose a granularity (clause; sentence; paragraph),
  3. create a similarity measure for that granularity (word overlap; multi-word overlap, perfect match),
  4. measure the similarity of each unit in the new to the most similar unit(s) in the gold standard,
  5. measure Recall and Precision.

  e.g., (Kupiec et al., 95).

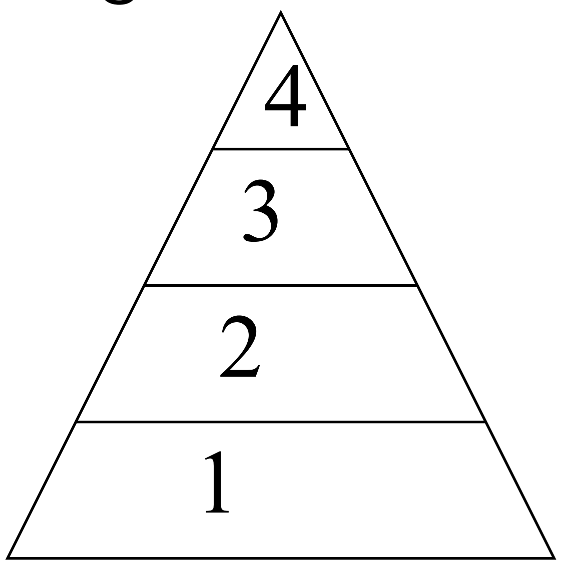
Evaluation

- Manual
  - Linguistic Quality (readability)
    - Grammaticality
    - Non-redundancy
    - Referential clarity
    - Focus
    - Structure
  - Five-point scale (1 very poor, 5 very good)
- Pyramid: Summarization Content Units (SCUs)
  - Annotate of a corpus of summaries and are not bigger than a clause.
- Automatic
  - Rouge
    - ROUGE 2
    - ROUGE SU4



Pyramid(1)

- The pyramid method is designed to address the observation: summaries from different humans always have partly overlapping content.
- The pyramid method includes a manual annotation method to represent Summary Content Units (SCUs) and to quantify the proportion of model summaries that express this content.
- All SCUs have a weight representing the number of models they occur in, thus from 1 to max_n, where max_n is the total number of models
    - There are very few SCUs expressed in all models (i.e., weight=max_n), and increasingly many SCUs at each lower weight, with the most SCUs at weight=1.



SCU example

A1 In 1998 two Libyans indicted in 1991 for the Lockerbie bombing were still in Libya.
B1 Two Libyans were indicted in 1991 for blowing up a Pan Am jumbo jet over Lockerbie, Scotland in 1988.
C1 Two Libyans, accused by the United States and Britain of bombing a New York bound Pan Am jet over Lockerbie, Scotland in 1988, killing 270 people, for 10 years were harbored by Libya who claimed the suspects could not get a fair trail in America or Britain.
D2 Two Libyan suspects were indicted in 1991.

SCU1 (w=4): two Libyans were officially accused of the Lockerbie bombing
A1 [two Libyans]1 [indicted]1
B1 [Two Libyans were indicted]1
C1 [Two Libyans,]1 [accused]1
D2 [Two Libyan suspects were indicted]1

ROUGE basics

ROUGE basics

- Rouge (Recall-Oriented Understudy for Gisting Evaluation)
  Recall-oriented, within-sentence word overlap with model(s)

ROUGE basics

- Rouge (Recall-Oriented Understudy for Gisting Evaluation) Recall-oriented, within-sentence word overlap with model(s)
- Models - no theoretical limit to number
  - compared system output to 4 models
  - compared manual summaries to 3 models

ROUGE basics

- Rouge (Recall-Oriented Understudy for Gisting Evaluation) Recall-oriented, within-sentence word overlap with model(s)
- Models - no theoretical limit to number
  - compared system output to 4 models
  - compared manual summaries to 3 models
- Using n-gram
  - Correlate reasonably with human coverage judgements
  - Not address summary discourse characteristics, and suffer from lack of text cohesion or coherence

ROUGE basics

- Rouge (Recall-Oriented Understudy for Gisting Evaluation) Recall-oriented, within-sentence word overlap with model(s)
- Models - no theoretical limit to number
  - compared system output to 4 models
  - compared manual summaries to 3 models
- Using n-gram
  - Correlate reasonably with human coverage judgements
  - Not address summary discourse characteristics, and suffer from lack of text cohesion or coherence
- ROUGE v1.2.1 measures
  - ROUGE-1,2,3,4: N-gram matching where N = 1,2,3,4
  - ROUGE-LCS: Longest common substring

ROUGE

$$R_n(X) = \frac{\sum_{j=1}^{h} \sum_{i \in N_n} \min(X_n(i), M_n(i,j))}{\sum_{j=1}^{h} \sum_{i \in N_n} M_n(i,j)}$$

- Where *Nn* represents the set of all *n*-grams and *i* is one member from *Nn*. *Xn(i)* is the number of times the n-gram *i* occurred in the summary and *Mn(i,j)* is the number of times the n-gram *i* ocurred in the *j*-th model reference(human) summary. There are totally *h* human summaries.

Questions (from Spärck Jones)

- Should we take the reader into account and how?
- “Similarly, the notion of a basic summary, i.e., one reflective of the source, makes hidden fact assumptions, for example that the subject knowledge of the output’s readers will be on a par with that of the readers for whom the source was intended. (p. 5)”
- Is the state of the art sufficiently mature to allow summarization from intermediate representations and still allow robust processing of domain independent material?

What should you know

What should you know

- How to use semi-supervised learning for information extraction and what is distant supervision

What should you know

- How to use semi-supervised learning for information extraction and what is distant supervision
- Textual Entailment / Natural Language Inference lets us understand whether some sentence is likely true based on a prior piece of text

What should you know

- How to use semi-supervised learning for information extraction and what is distant supervision
- Textual Entailment / Natural Language Inference lets us understand whether some sentence is likely true based on a prior piece of text
- The basics of Question Answering and how knowledge bases are used

What should you know

- How to use semi-supervised learning for information extraction and what is distant supervision
- Textual Entailment / Natural Language Inference lets us understand whether some sentence is likely true based on a prior piece of text
- The basics of Question Answering and how knowledge bases are used
- What are the difference challenges in text summarization