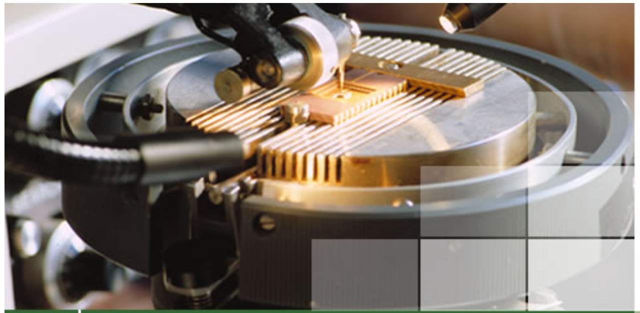
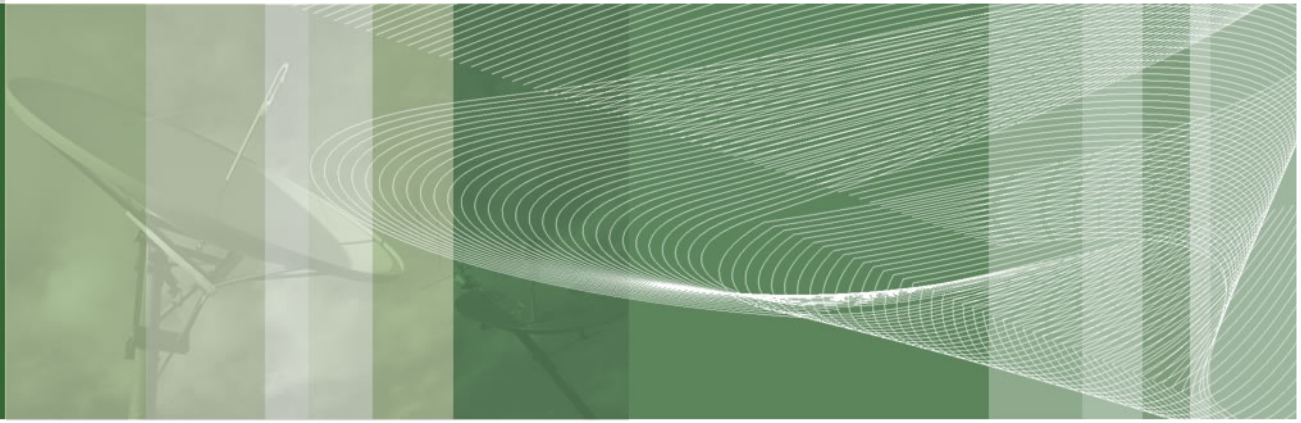


POLITECNICO DI MILANO

Dipartimento di
Elettronica e Informazione



Search Computing

Natural Language Processing for Information Retrieval

Silvia Quarteroni

quarteroni@elet.polimi.it

Outline

- Motivations
- Part I: Basic Concepts
- Part II: Applications
- Conclusions

MOTIVATIONS

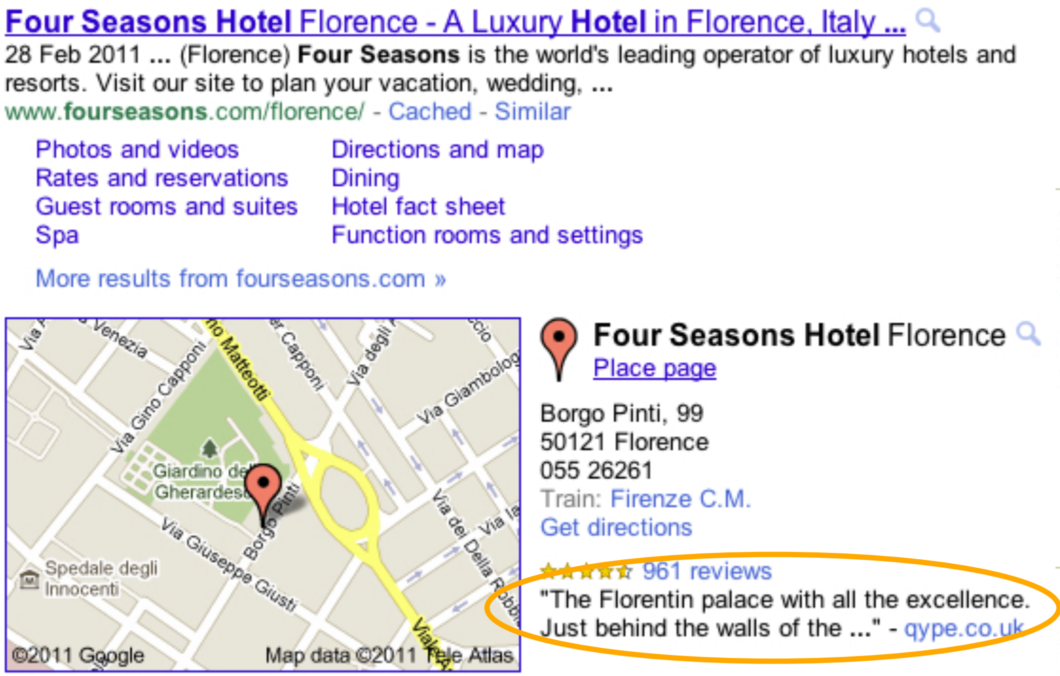
Why Natural Language Processing ?

- Huge amounts of data
  - Internet = at least 20 billion pages
  - Intranet
- Applications for processing large amounts of texts require NLP expertise

- Classify text into categories
- Index and search large texts
- Automatic translation
- Speech understanding
- Information extraction
- Automatic summarization
- Question answering
- Knowledge acquisition
- Text generation
- Dialog management

Linguistic data is ubiquitous

- Most of the information around companies & the Web comes in human languages – not traditional DB stuff!
  - reports, customer email,
  - web pages, sound, video,
  - opinions, feedback



Four Seasons Hotel Florence - A Luxury **Hotel** in Florence, Italy ...

28 Feb 2011 ... (Florence) **Four Seasons** is the world's leading operator of luxury hotels and resorts. Visit our site to plan your vacation, wedding, ...

www.**fourseasons**.com/florence/ - Cached - Similar

Photos and videos	Directions and map
Rates and reservations	Dining
Guest rooms and suites	Hotel fact sheet
Spa	Function rooms and settings

More results from fourseasons.com »

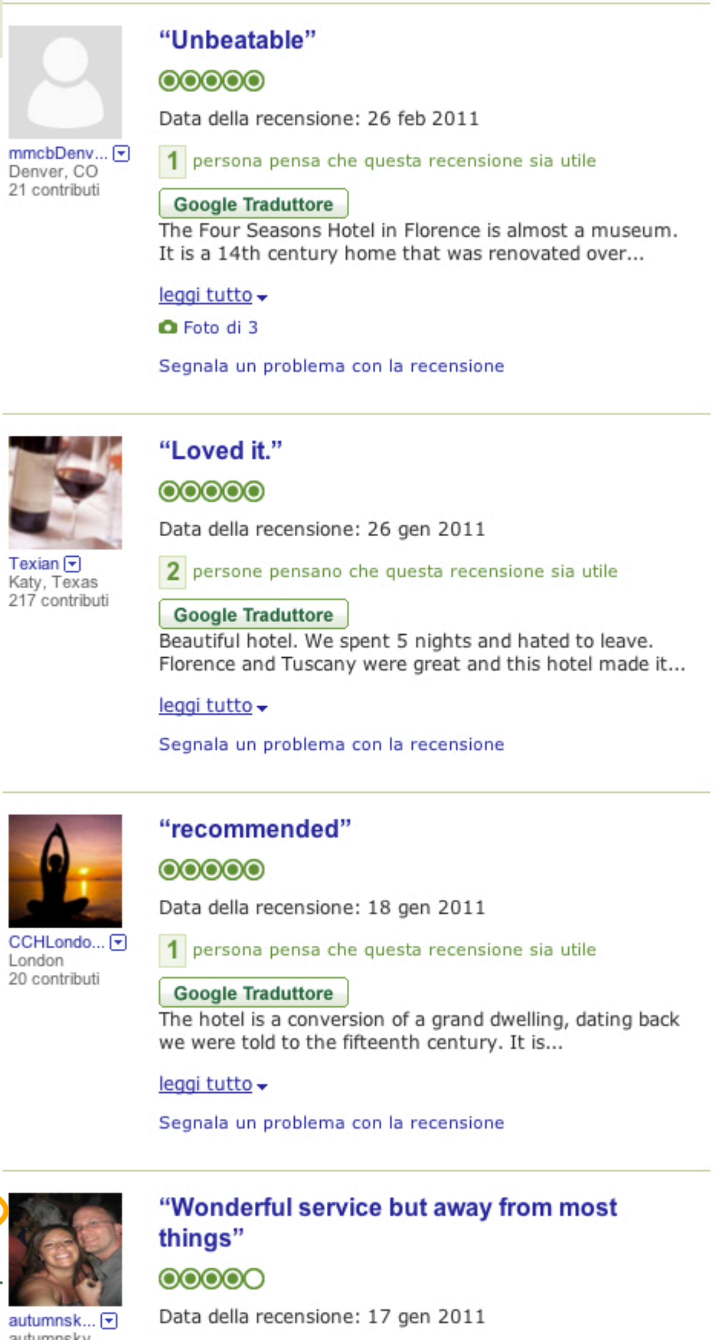
Four Seasons Hotel Florence

Place page

Borgo Pinti, 99
50121 Florence
055 26261
Train: Firenze C.M.
Get directions

★★★★☆ 961 reviews

"The Florentin palace with all the excellence. Just behind the walls of the ..." - qype.co.uk



"Unbeatable"

●●●●●

Data della recensione: 26 feb 2011

mmcbDenv... Denver, CO 21 contributi

1 persona pensa che questa recensione sia utile

Google Traduttore

The Four Seasons Hotel in Florence is almost a museum. It is a 14th century home that was renovated over...

leggi tutto

Foto di 3

Segnala un problema con la recensione

"Loved it."

●●●●●

Data della recensione: 26 gen 2011

Texian Katy, Texas 217 contributi

2 persone pensano che questa recensione sia utile

Google Traduttore

Beautiful hotel. We spent 5 nights and hated to leave. Florence and Tuscany were great and this hotel made it...

leggi tutto

Segnala un problema con la recensione

"recommended"

●●●●●

Data della recensione: 18 gen 2011

CCHLondo... London 20 contributi

1 persona pensa che questa recensione sia utile

Google Traduttore

The hotel is a conversion of a grand dwelling, dating back we were told to the fifteenth century. It is...

leggi tutto

Segnala un problema con la recensione

"Wonderful service but away from most things"

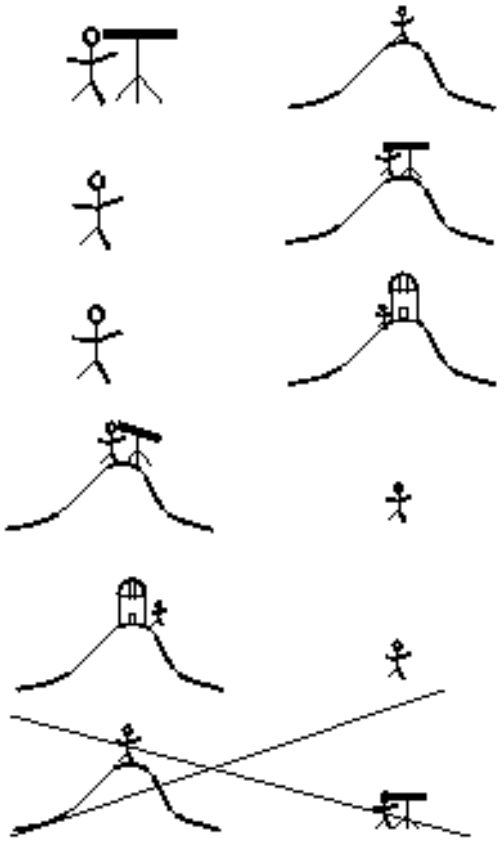
●●●●○

Data della recensione: 17 gen 2011

autumnsk... autumnsky

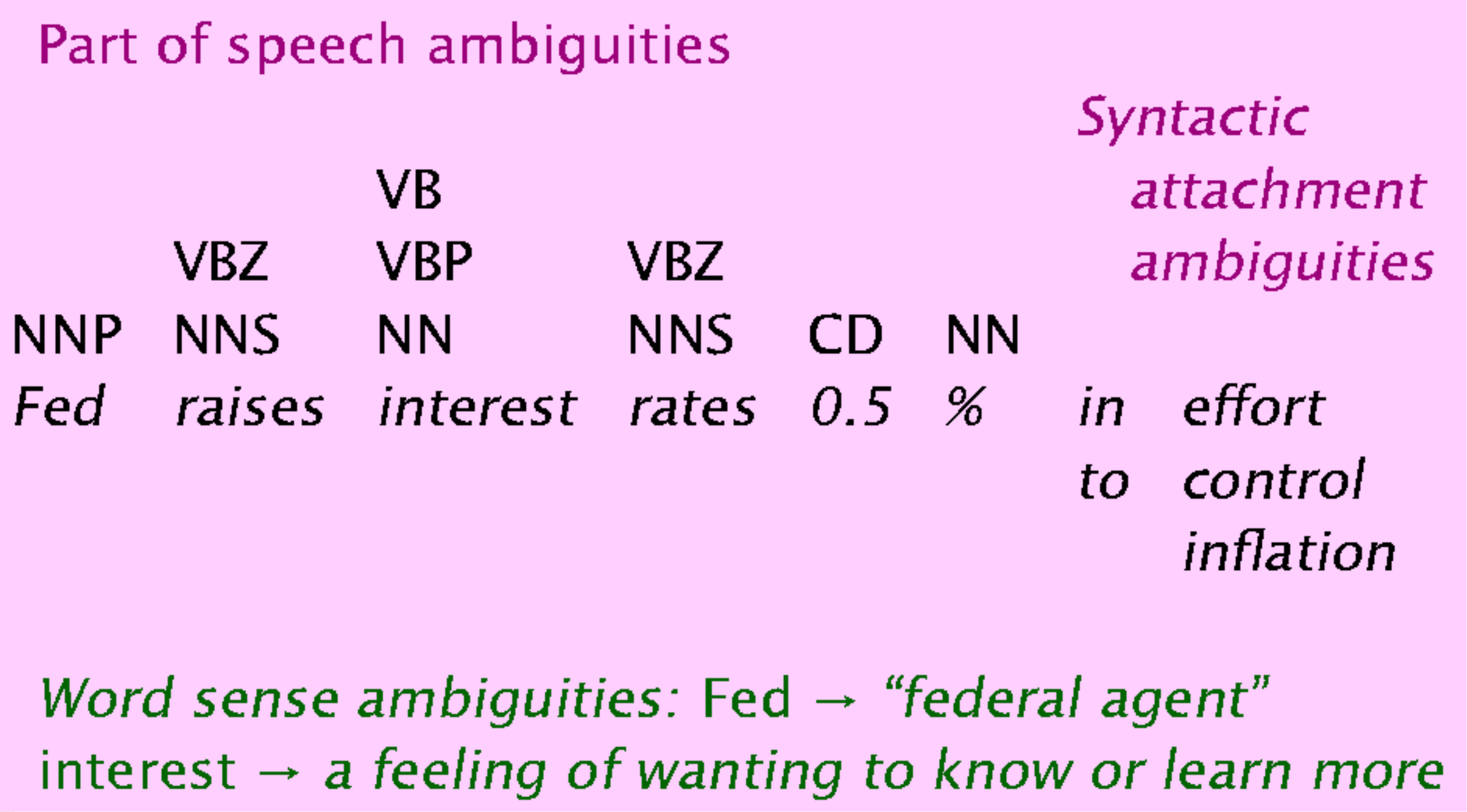
Why is Natural Language Understanding difficult?

- *Ambiguity* is the primary difference between natural language (NL) and computer languages (CLs)
  - CLs are designed by grammars that produce a unique parse for each sentence in the language
- Examples of ambiguous NL wordings
  - I saw the man on the hill with a telescope.
  - I saw the Grand Canyon flying to LA.
  - Time flies like an arrow.
  - Fruit flies like banana.



Resolving ambiguity

- The hidden structure of language is highly ambiguous at different levels: lexical, syntactic, semantic



Part of speech ambiguities

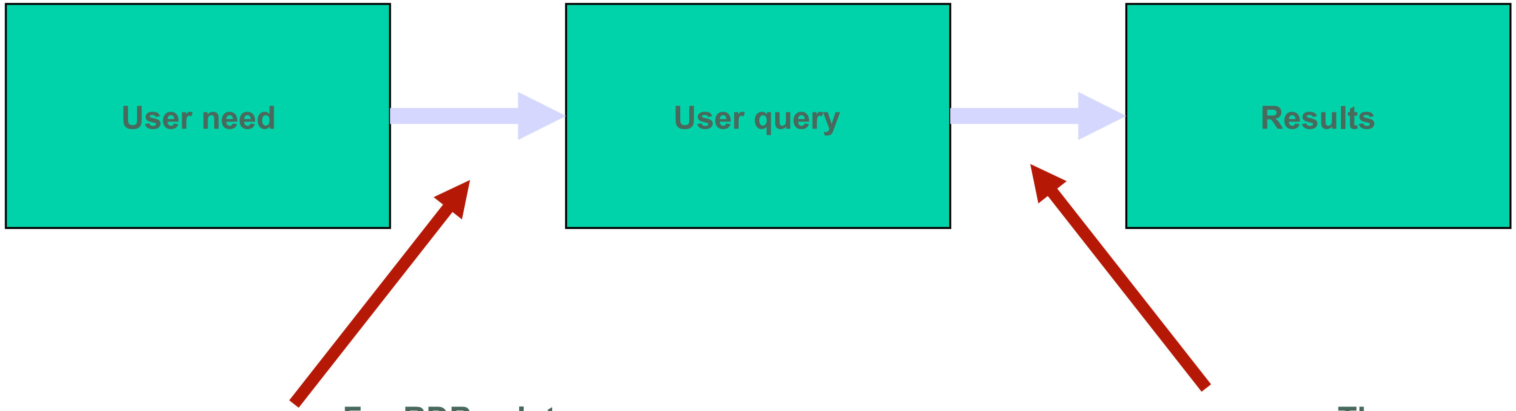
Syntactic attachment ambiguities

		VB					
	VBZ	VBP	VBZ				
NNP	NNS	NN	NNS	CD	NN		
Fed	*raises*	*interest*	*rates*	*0.5*	*%*	*in*	*effort*
						to	*control*
							inflation

Word sense ambiguities: Fed → *"federal agent"*

interest → *a feeling of wanting to know or learn more*

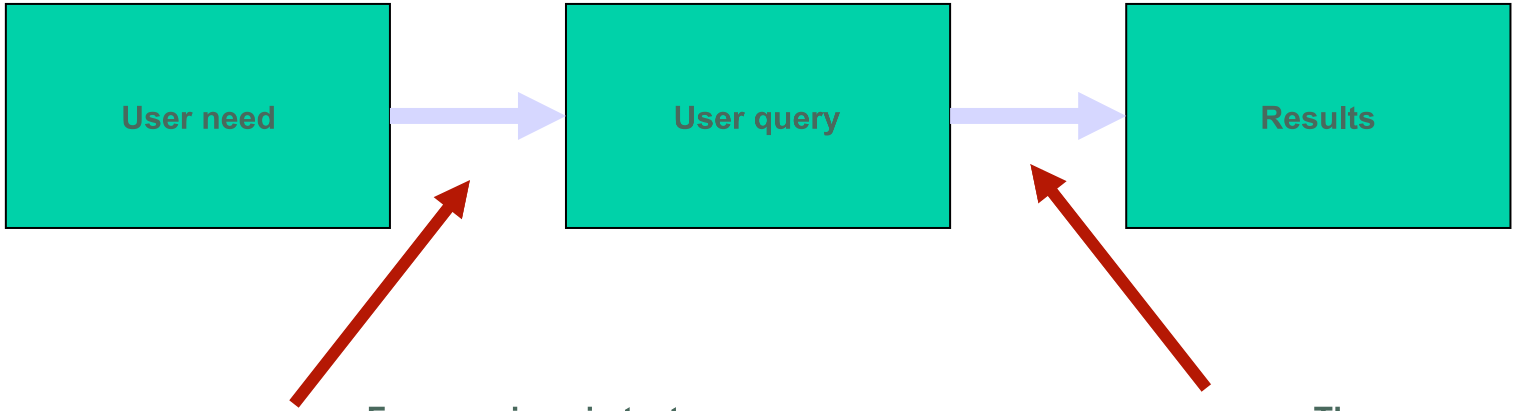
Translating user needs: DBs



For RDB, a lot of people know how to do this correctly, using SQL or a GUI tool

The answers coming out here will then be precisely what the user wanted

Translating user needs: IR

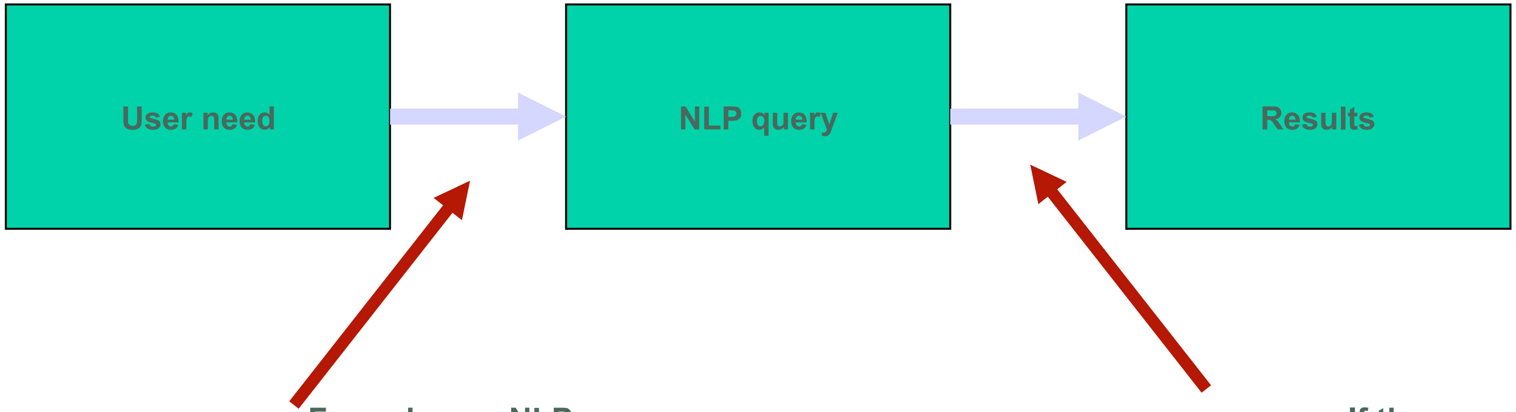


For meanings in text,
no IR-style query
gives one exactly
what one wants;
it only hints at it

The answers
coming out may
be roughly what
was wanted, or
can be refined

Sometimes!

Translating user needs: NLP



User need

NLP query

Results

For a deeper NLP analysis system, the system subtly translates the user’s language

If the answers coming back aren’t what was wanted, the user frequently has *no idea* how to fix the problem

Risky!

Part I

BASIC CONCEPTS

Summary

- Definitions
- Core technologies
- Methods
- Evaluation

Natural Language Processing

- NLP is the branch of computer science focused on developing systems that allow computers to communicate with people using everyday language.
- Also called Computational Linguistics
  - More on the linguistically motivated side of the problem
  - Also deals with computational methods to understand human language

NLP and IR

- Several NLP methods and applications are strictly related to IR
  - **Text Categorization**
  - **Question Answering**
  - Semantic Search
  - Opinion Mining
  - Spoken Language Understanding
  - Clustering
  - …

A brief history of NLP: 1950s-1980s

- 1950s: the Turing test
  - Soon enough, machines would be mistaken for humans
- 1960s-1970s:
  - Thanks to simple pattern-matching rules, ELIZA the chat-bot [Weizenbaum,1966] was able to converse in natural language
  - In the SHRDLU world [Winograd,1971], a mechanical hand would receive commands in NL to move blocks around
  - “Conceptual ontologies” to represent knowledge in restricted domains, rule-based approaches to NL understanding [Schank & Abelson, 1977]

A brief history of NLP: 1980s-1990s

- 1980s-90s: Machine Learning & statistical models emerge
  - Decision trees, Support Vector Machines [Joachims,1998], Hidden Markov Models, …
  - Syntactic parsers, Named Entity recognizers trained on large datasets [Finkel et al,2005]
  - Large news/medical corpora made available to test algorithms using deep NL features
    - NYT, WSJ
    - Medline, PubMed

A brief history of NLP: 1990s-2000s

- 1990s-2000s:
  - great NLP evaluation campaigns
    - TREC (trec.nist.gov), CLEF (clef-campaign.org)
    - challenging tasks such as word sense disambiguation, summarization, question answering, machine translation
  - Deeper NLP:
    - semantic role labeling [Carreras & Marquez, 2005] shifts analysis from syntax to semantics: predicate-argument relations
  - Industrial mobile NL technologies (automatic speech recognition/understanding) push NLP towards more and more robustness
    - AT&T’s *How May I Help You?* [Gorin et al.,1997]

A brief history of NLP: today

- Today:
  - Still a lot of Machine Learning: discriminative methods such as SVMs, Conditional Random Fields
  - Challenging problems: answering complex questions (Watson wins *Jeopardy!*), machine translation
  - A lot of effort on non-text: *speech* understanding now makes it possible to have you speak your Google search

Summary

- Definitions
- Core technologies
- Methods
- Evaluation
- Conclusions

Levels of understanding: Syntax, Semantics, Pragmatics

- **Syntax** concerns the proper ordering of words and its effect on meaning.
  - *The dog bit the boy != The boy bit the dog.*
- **Semantics** concerns the (literal) meaning of words, phrases, and sentences.
  - "plant": a photosynthetic organism, a manufacturing facility, the act of sowing
- **Pragmatics** concerns the overall communicative and social context and its effect on interpretation.
  - *Remove the kernels from the cherries and throw them away*

Syntactic Tasks

- Word Segmentation
- Morphological Analysis
- Part of Speech Tagging
- Shallow Parsing
- Deep Syntactic Parsing

Morphological Analysis

- ***Morphology*** is the field of linguistics that studies the internal structure of words. (Wikipedia)
- A ***morpheme*** is the smallest linguistic unit that has semantic meaning (Wikipedia)
  - e.g. “carry”, “pre”, “ed”, “ly”, “s”
- Morphological analysis is the task of segmenting a word into its morphemes:
  - carried ⇒ carry + ed (past tense)
  - independently ⇒ in + (depend + ent) + ly
  - Googlers ⇒ (Google + er) + s (plural)
  - unlockable ⇒ un + (lock + able) ?

    ⇒ (un + lock) + able ?

Part Of Speech (POS) Tagging

- Annotate each word in a sentence with a part-of-speech.

I Pro ate V the Det spaghetti N with Prep meatballs N.

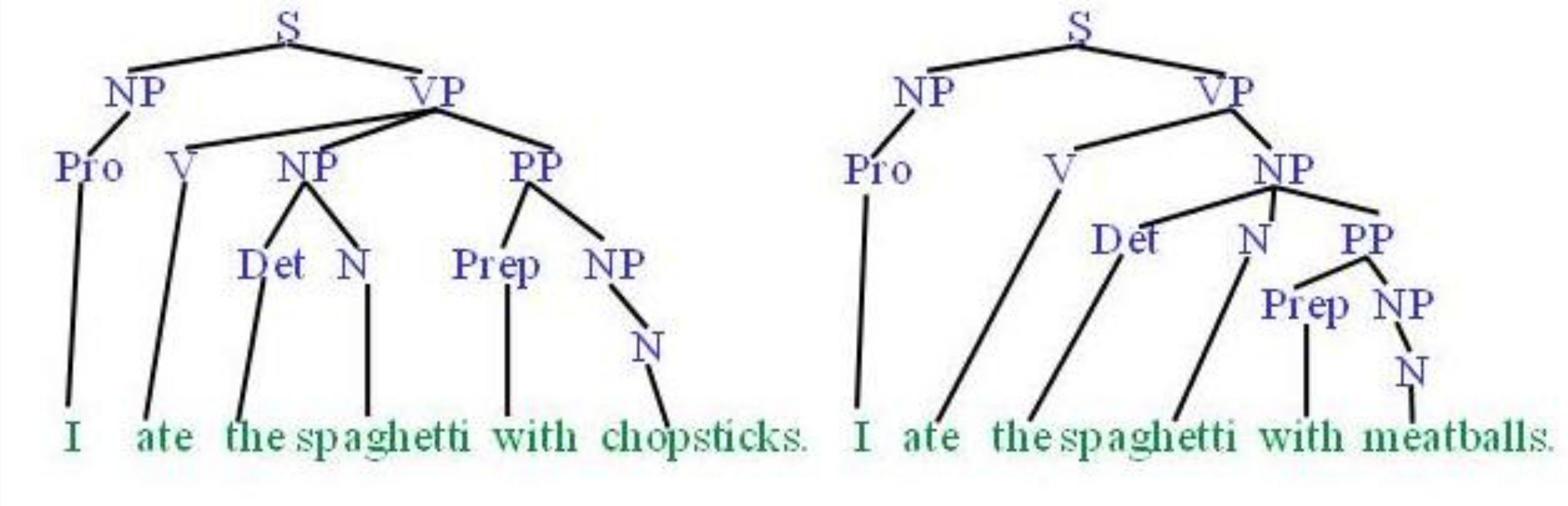
- Useful for subsequent tasks such as syntactic parsing and word sense disambiguation
- “Classic” approach: decision trees [Schmid’94]
- POS tagging is considered to be a “solved problem”, with > 90% accuracy
- POS taggers exist for most languages, even least researched ones

Phrase Chunking (aka Shallow Parsing)

- Find all non-recursive noun phrases (NPs) and verb phrases (VPs) in a sentence.
  - [NP I] [VP ate] [NP the spaghetti] [PP with] [NP meatballs].
  - [NP He] [VP reckons] [NP the current account deficit] [VP will narrow] [PP to] [NP only # 1.8 billion] [PP in] [NP September]
- Requires a segmentation part (separate phrases from each other) and a tagging part (tag separated phrases)
- Many packages exist for chunking, cf opennlp at Stanford (opennlp.sourceforge.net)

Syntactic Parsing

- Produce the correct syntactic parse tree for a sentence.


S
NP
VP
Pro
V
NP
PP
Det
N
Prep
NP
N
I ate the spaghetti with chopsticks.
S
NP
VP
Pro
V
NP
Det
N
PP
Prep
NP
N
I ate the spaghetti with meatballs.

Semantic Tasks

- Word Sense Disambiguation
- Semantic Role Labeling
- Recognizing Textual Entailment

Word Sense Disambiguation (WSD)

- Words in natural language usually have a fair number of different possible meanings.
  - Ellen has a strong interest in computational linguistics.
  - Ellen pays a large amount of interest on her credit card.
- For many tasks (question answering, translation), the proper sense of each ambiguous word in a sentence must be determined.

Semantic Role Labeling (SRL)

- For each clause, determine the semantic role played by each noun phrase that is an argument to the verb.

  agent patient source destination instrument

  - John drove Mary from Austin to Dallas in his Toyota Prius.
  - The hammer broke the window.

- Also referred to a "case role analysis," "thematic analysis," and "shallow semantic parsing"

Textual Entailment

- Determine whether one natural language sentence entails (implies) another under an ordinary interpretation. Example from PASCAL RTE challenge:

TEXT	HYPOTHESIS	ENTAILMENT
Eyeing the huge market potential, currently led by Google, Yahoo took over search company Overture Services Inc. last year.	*Yahoo bought Overture.*	TRUE
Microsoft's rival Sun Microsystems Inc. bought Star Office last month and plans to boost its development as a Web-based device running over the Net on personal computers and Internet appliances.	*Microsoft bought Star Office.*	FALSE
The National Institute for Psychobiology in Israel was established in May 1971 as the Israel Center for Psychobiology by Prof. Joel.	*Israel was established in May 1971.*	FALSE
Since its formation in 1948, Israel fought many wars with neighboring Arab countries.	*Israel was established in 1948.*	TRUE

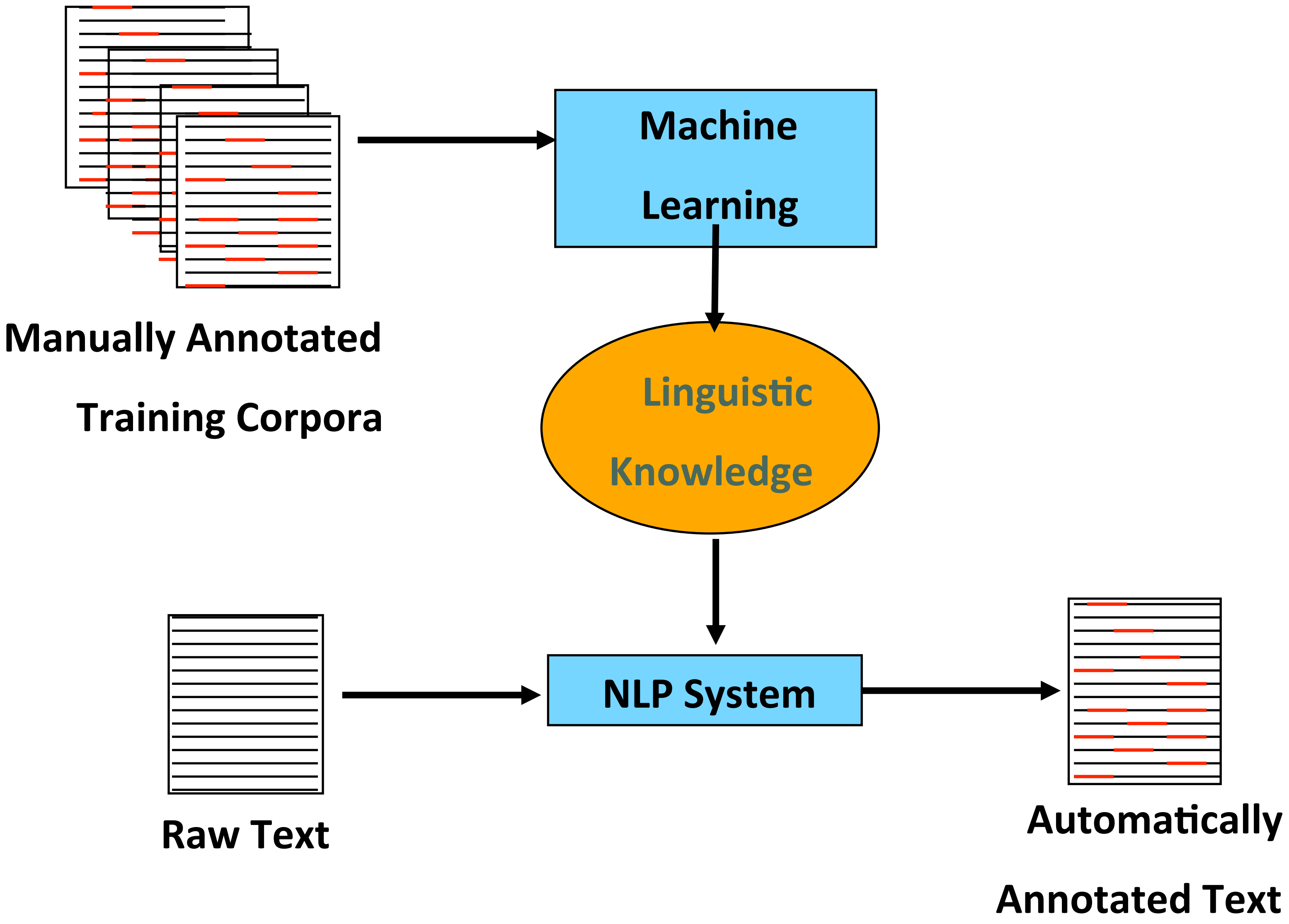
Pragmatic/Discourse tasks

- Anaphora: an instance of an expression referring to another
- Co-reference occurs when multiple expressions in a sentence or document have the same referent (i.e. refer to the same phrase).
- Anaphora/co-reference resolution consists in determining which phrases in a document refer to the same underlying entity.
  - John put the carrot on the plate and ate it.
  - Bush started the war in Iraq. But the president needed the consent of Congress.
- Ellipsis is the omission or suppression of parts of words or sentences when these can be inferred from the context
  1. Wise men talk because they have something to say; fools because they have to say something (Plato)
  2. Wise men talk because they have something to say; fools talk because they have to say something (Plato)

Summary

- Definitions
- Core technologies
- Methods
- Evaluation
- Conclusions

Machine Learning Approach



Advantages of the Learning Approach

- Larger and larger amounts of text are available.
- Annotating corpora is easier and requires less expertise than manual knowledge engineering.
- Learning algorithms have progressed to be able to handle large amounts of data and produce accurate probabilistic knowledge.
- The probabilistic knowledge acquired allows robust processing that handles linguistic regularities as well as exceptions.

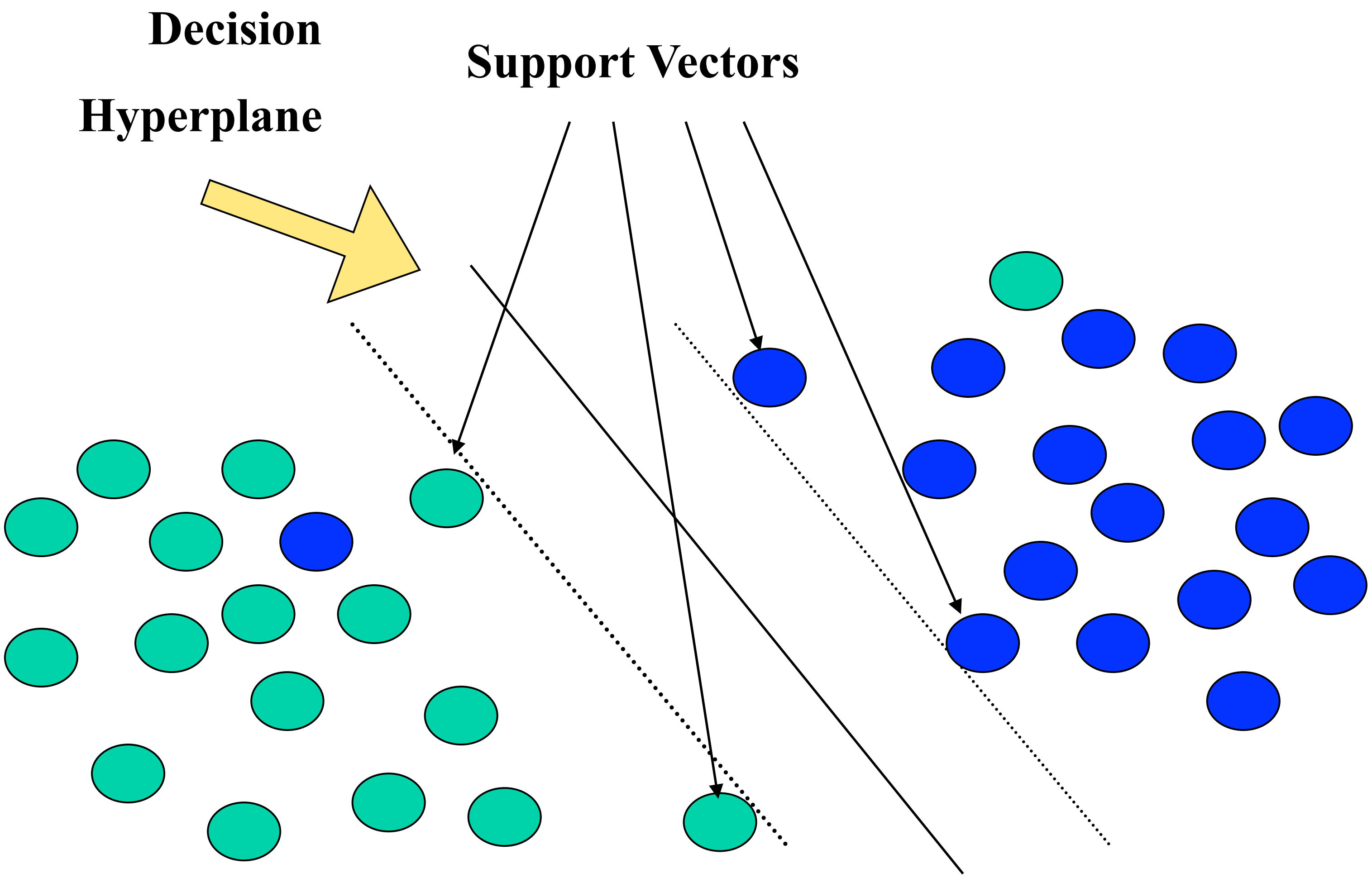
The Importance of Probability

- Unlikely interpretations of words can combine to generate spurious ambiguity:
  - "Time flies like an arrow" might be interpreted as:
    - Insects of a variety called "time flies" are fond of a particular arrow.
    - A command to record insect speed in the manner that an arrow would.
- Some combinations of words are more likely than others:
  - "wreck a nice beach" vs. "recognize speech"
  - "vice president Gore" vs. "dice precedent core"
  - "Let us pray" vs. "Lettuce spray"
- Statistical methods compute the most likely interpretation by combining probabilistic evidence from a variety of uncertain knowledge sources.

Machine Learning for NLP

- Many NLP problems can be reduced to a *supervised classification* task
  - Given a set of training examples (with known class label), learn a model
  - Use the model on a set of testing examples to guess their label
- Examples:
  - Text categorization
  - Query classification
  - Question/candidate answer classification
  - Opinion polarity identification

Support Vector Machines: intuition



Support Vector Machines

- Idea: learn a decision hyperplane *H* separating instances of two classes (POS vs NEG)
  - *H* is described by $\mathbf{w} \cdot \mathbf{x} - b = 0,$
  - $\mathbf{w} \cdot \mathbf{x}_i - b \geq 1 \qquad \text{for } \mathbf{x}_i$ in POS,
  - $\mathbf{w} \cdot \mathbf{x}_i - b \leq -1 \qquad \text{for } \mathbf{x}_i$ in NEG.
  - We can rewrite this as

    $$y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1, \quad \text{for all } 1 \leq i \leq n. \qquad (1)$$
- The weight vector **w** and the *b* constant are learned by examining all the **x** instances in the training data
- We find the optimal **w** and b by minimizing (1)
- This yields $\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x_i}$ (α_i are the Lagrange multipliers)
- Each $\mathbf{x_i}$ corresponding to an α_i is a Support Vector (SV)
- *b* is usually computed as $b = \frac{1}{N_{SV}} \sum_{i=1}^{N_{SV}} (\mathbf{w} \cdot \mathbf{x_i} - y_i)$

Summary

- Definitions
- Core technologies
- Methods
- Evaluation

Evaluating NLP systems

- Each task has
  - specific criteria
    - Most widely adopted are Precision/Recall variations
  - specific corpora/test collections
    - Generally, news/medical corpora with different levels of annotation

Standard test collections

- TREC - National Institute of Standards and Testing (trec.nist.gov) has run large NLP benchmarks for many years
  - TREC-QA, TREC-entity, TREC-med
  - TREC-blog, TREC-web, …
- CLEF – for cross language evaluation (clef-campaign.org)
  - Intellectual property
  - Multi-lingual Question Answering
  - Search for Entities on the Web, …
- NTCIR – east Asian languages and cross-language IR (ntcir.nii.ac.jp)
  - IR
  - Question Answering
  - Cross-lingual IR

Precision and Recall

- Precision (P): fraction of returned items that are correct
  - True Positives/ (True Positives + False Positives)
  - "degree of correctness" of the system
  - Does not consider the total number of items
- Recall (R): fraction of items to return that are effectively returned
  - (True Positives)/(True Positives + False Negatives)
  - "degree of completeness" of the system

F-measure

- F-measure: combined measure that assesses the tradeoff between precision and recall (weighted harmonic mean):

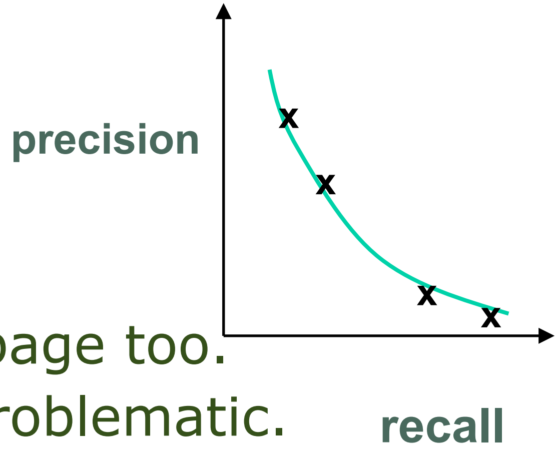
$$F = \frac{1}{\alpha\frac{1}{P} + (1-\alpha)\frac{1}{R}} = \frac{(\beta^2+1)PR}{\beta^2 P + R}$$

$$\beta^2 = \frac{1-\alpha}{\alpha}$$

  - Values of β<1 emphasize precision
  - Values of β>1 emphasize recall

- Most frequently used: balanced F-measure
  - β = 1 -> F = 2PR/(P+R)

Recall vs. Precision



- High recall:
  - You get all the right answers, but garbage too.
  - Good when incorrect results are not problematic.
  - More common from automatic systems.
- High precision:
  - When all returned answers must be correct.
  - Good when missing results are not problematic.
  - More common from hand-built systems.

Assessing ranked results

- **Mean Reciprocal Rank (MRR)** is a statistics for evaluating any process producing a list of possible responses to a query, ranked by probability of correctness.
- The reciprocal rank of a query response is the multiplicative inverse of the rank of the first correct answer.
- The Mean Reciprocal Rank is the average of reciprocal ranks of results for a sample of queries

$$\text{MRR} = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{\text{rank}_i}.$$

RANK	CASE A	CASE B
1	CORRECT	WRONG
2	WRONG	CORRECT
3	WRONG	CORRECT
MRR@3	1/3 = 0.33	(1/2 + 1/3)/3 = 0.25

Part II

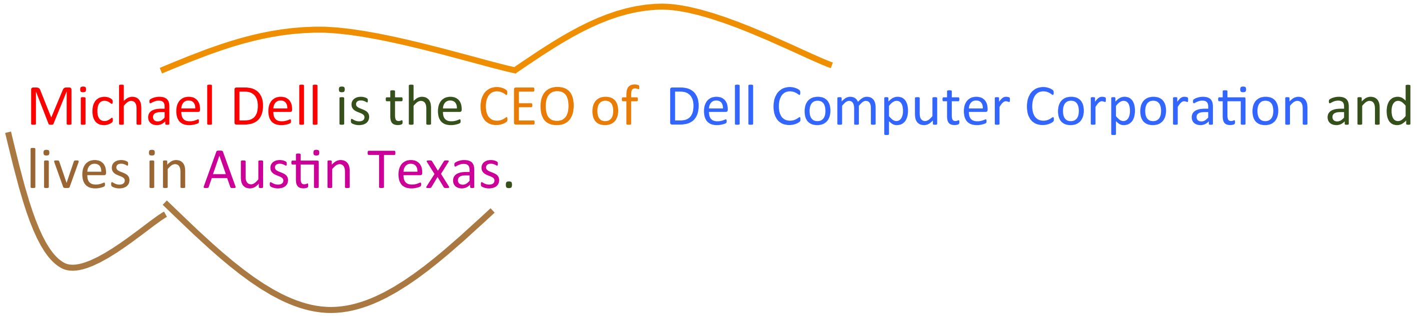
APPLICATIONS

Applications

- Information Extraction
- Machine Translation
- Summarization
- Spoken Language Understanding
- Opinion Mining/Sentiment Analysis
- Text categorization
- NLP on Mobile devices
- Question Answering

Information Extraction (IE)

- Identify phrases in language that refer to specific types of entities and relations in text.
  - Heavily based on **Named Entity Recognition**, identifying names of people, places, organizations, etc.
    - Michael Dell is the CEO of Dell Computer Corporation and lives in Austin Texas.
  - Also involves **Relation Extraction,** identifying specific relations between entities [Wong et al.'09, Zelenko et al'03].



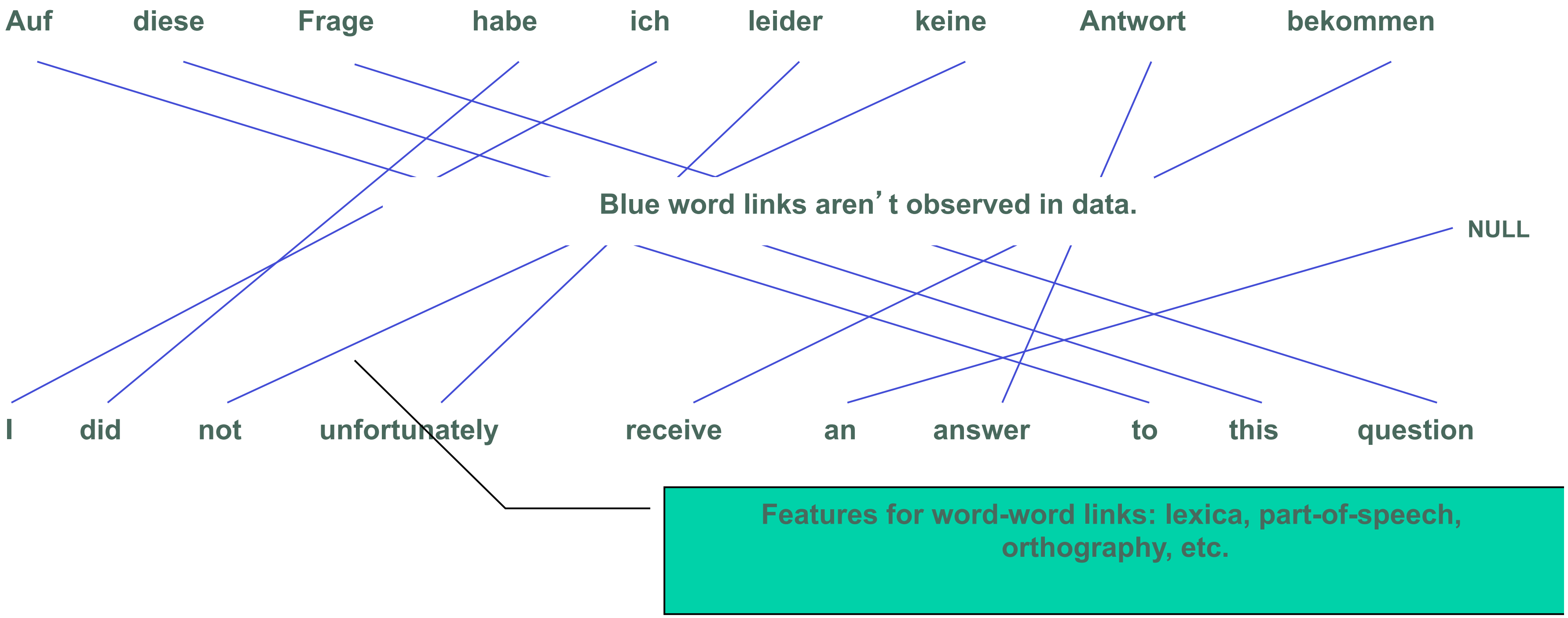
- Michael Dell is the CEO of Dell Computer Corporation and lives in Austin Texas.

Automatic Summarization

- Produce a short summary of a textual document
  - Article: *With a split decision in the final two primaries and a flurry of superdelegate endorsements, Sen. Barack Obama sealed the Democratic presidential nomination last night after a grueling and history-making campaign against Sen. Hillary Rodham Clinton that will make him the first African American to head a major-party ticket. Before a chanting and cheering audience in St. Paul, Minn., the first-term senator from Illinois savored what once seemed an unlikely outcome to the Democratic race with a nod to the marathon that was ending and to what will be another hard-fought battle, against Sen. John McCain, the presumptive Republican nominee….*
  - Summary: *Senator Barack Obama was declared the presumptive Democratic presidential nominee.*
- Approaches:
  - Supervised: as in key-phrase extraction, *extractive summarization* selects sentences which mostly represent text based on e.g. TF*IDF
  - Unsupervised: find "centroid" sentence s*, i.e. mean word vector of all the sentences in the document. Rank other sentences by similarity to s*

Machine Translation

- Translate a sentence from one natural language to another.
  - Find optimal translation $\tilde{e} = arg\max_{e \in e^*} p(e|f) = arg\max_{e \in e^*} p(f|e)p(e)$
- Current methods are *statistical*: learn a language model for source & target language, then align **phrases** based on learned criteria



Auf diese Frage habe ich leider keine Antwort bekommen

Blue word links aren’t observed in data.

NULL

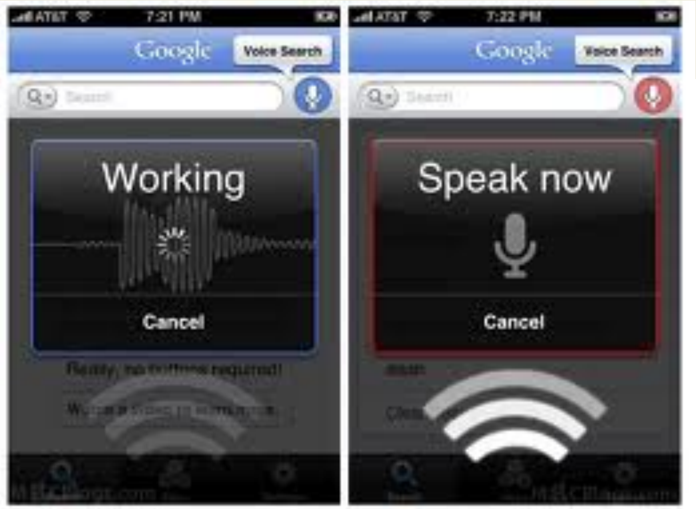
I did not unfortunately receive an answer to this question

Features for word-word links: lexica, part-of-speech, orthography, etc.

Opinion Mining/Sentiment Analysis

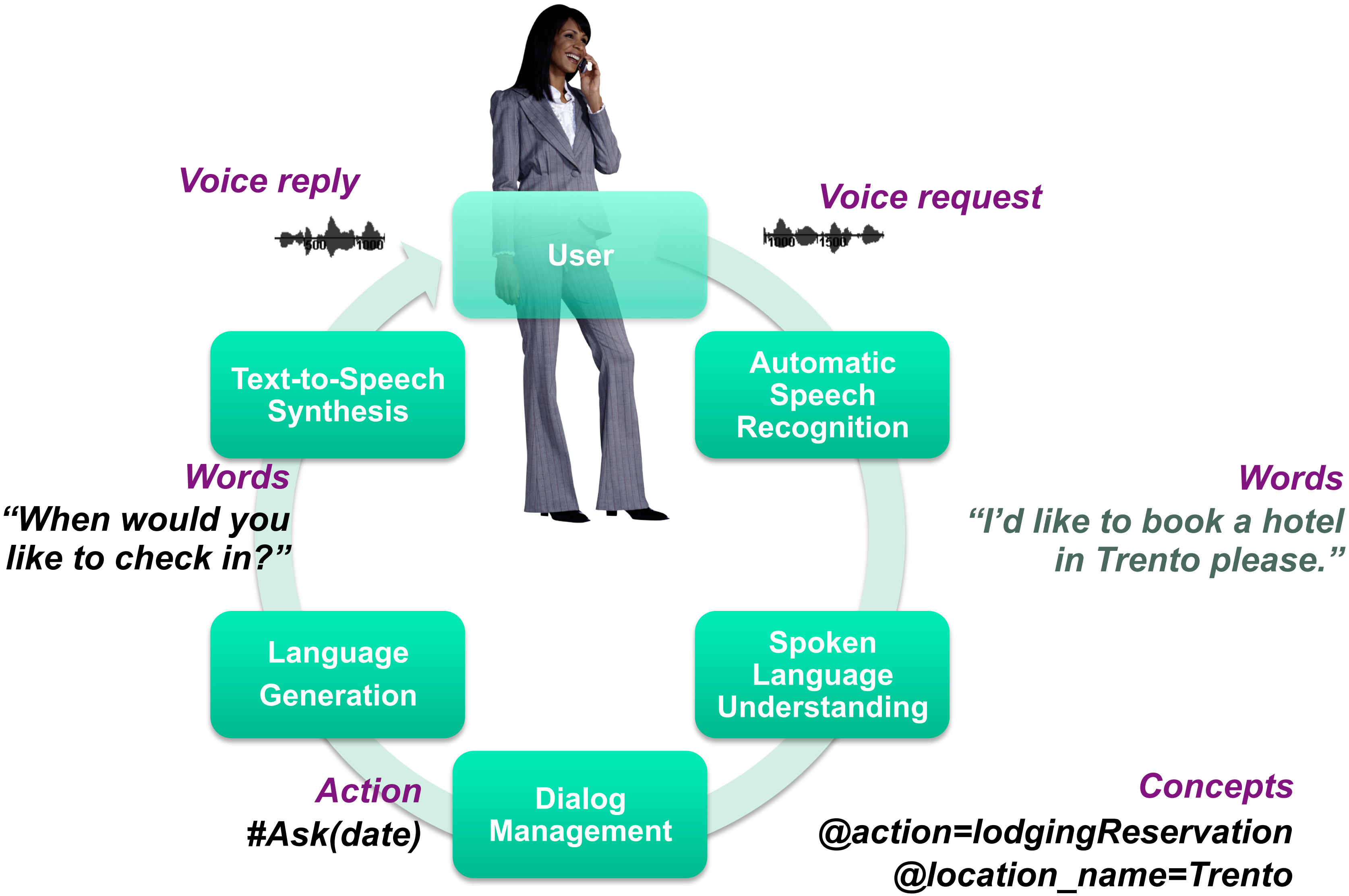
- Public subjective data on the web
  - Twitter, TripAdvisor, reviews…
- The signal to noise ratio is very low, but there's still lots of good information there
- Some of it has commercial value
  - What problems have users had with your product?
  - Why did people end up buying product X rather than your product Y?
- Some of it is time sensitive
  - Rumors on chat rooms can affect stock price
    - Regardless of whether they are factual or not
- Methods: largely bag-of-words based binary classifier [Turney'02]

NLP on Mobile



- With a big monitor, humans can scan for the right information
- On a mobile device, different interaction modalities may come handy
- Speech recognition and understanding have been a hot topic for roughly 30 years now
  - Today, you can speak to Google on your mobile to formulate a query
  - Several companies/institutions have automatic helpdesks providing info and troubleshooting users via phone
  - Ever tried Siri?

The Spoken Dialog System “loop”



Voice reply

Voice request

User

Automatic Speech Recognition

Words

“I’d like to book a hotel in Trento please.”

Spoken Language Understanding

Concepts

@action=lodgingReservation
@location_name=Trento

Dialog Management

Action

#Ask(date)

Language Generation

Words

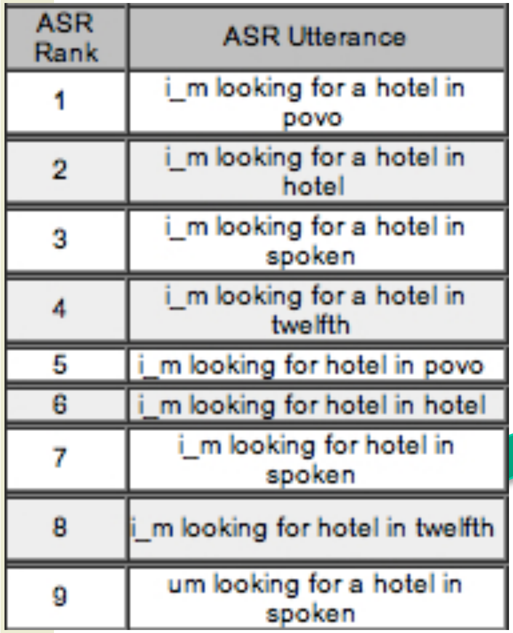
“When would you like to check in?”

Text-to-Speech Synthesis

The SDS "loop"

- Speech processing comes in a "loop"
- Each step is characterized by dedicated methods, largely statistical
  - Estimate probability that acoustic signal corresponds to a given word sequence (language model)
  - Classify utterance into problem class (SVM, CRF)
  - Plan next conversation step to maximize task success (Reinforcement Learning)

Models for robustness in ASR and SLU: Interpretation confidence



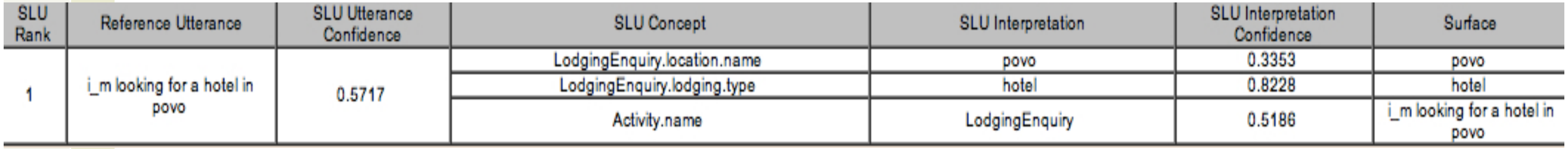
ASR Rank	ASR Utterance
1	i_m looking for a hotel in povo
2	i_m looking for a hotel in hotel
3	i_m looking for a hotel in spoken
4	i_m looking for a hotel in twelfth
5	i_m looking for hotel in povo
6	i_m looking for hotel in hotel
7	i_m looking for hotel in spoken
8	i_m looking for hotel in twelfth
9	um looking for a hotel in spoken

- Confidence = log P(correct)
  - A self-estimation of correctness
- Concept confidence:

  *f(*confidences of concept words*)*
- SLU interpretation confidence:

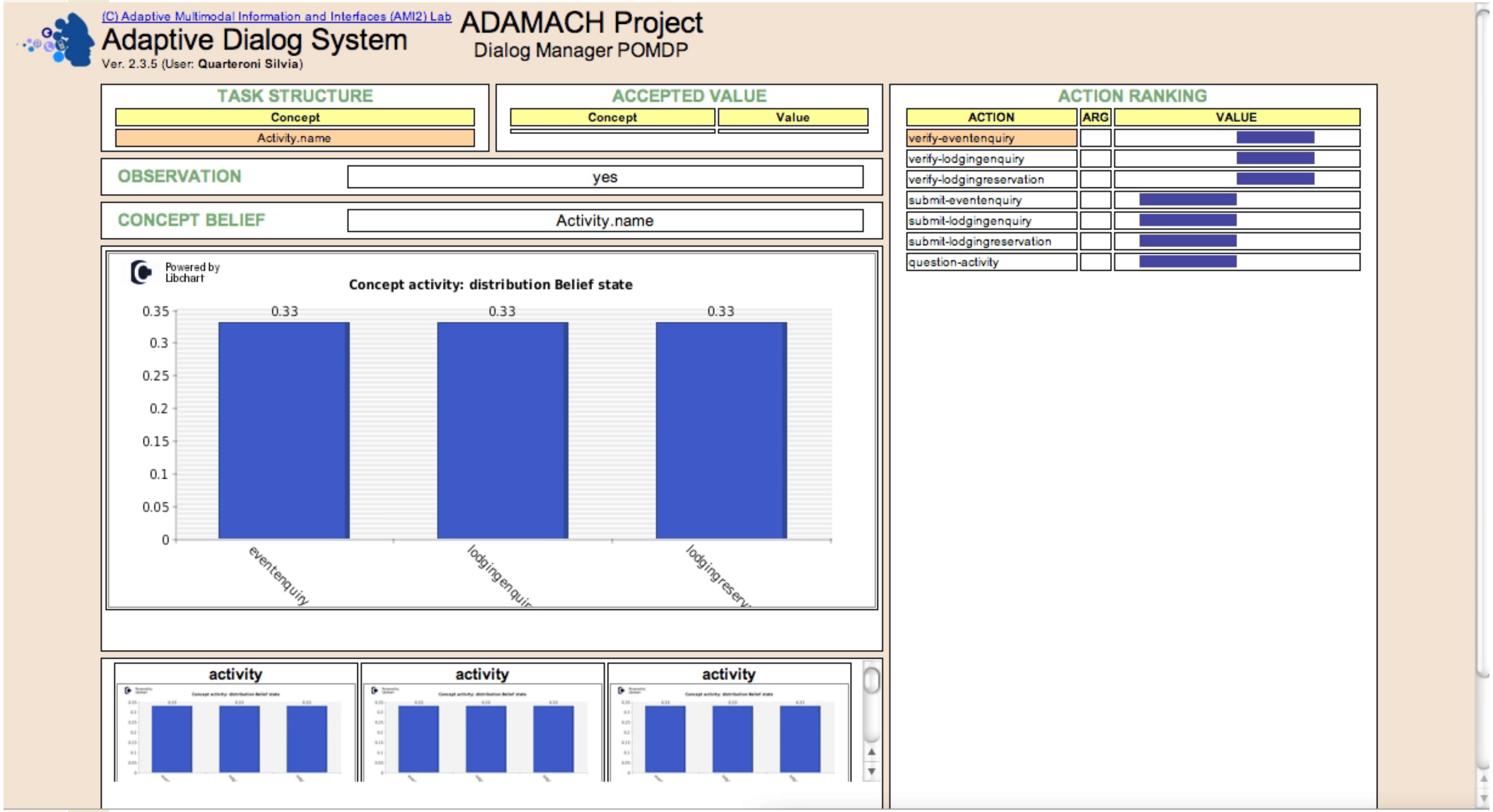
  *f(*confidence of concepts*)*

  2 top interpretations are returned to Dialog Manager



SLU Rank	Reference Utterance	SLU Utterance Confidence	SLU Concept	SLU Interpretation	SLU Interpretation Confidence	Surface
1	i_m looking for a hotel in povo	0.5717	LodgingEnquiry.location.name	povo	0.3353	povo
			LodgingEnquiry.lodging.type	hotel	0.8228	hotel
			Activity.name	LodgingEnquiry	0.5186	i_m looking for a hotel in povo

Demo: A Statistical Dialog System



(C) Adaptive Multimodal Information and Interfaces (AMI2) Lab

Adaptive Dialog System

Ver. 2.3.5 (User: **Quarteroni Silvia**)

ADAMACH Project

Dialog Manager POMDP

TASK STRUCTURE

Concept
Activity.name

ACCEPTED VALUE

Concept	Value

OBSERVATION: yes

CONCEPT BELIEF: Activity.name

ACTION RANKING

ACTION	ARG	VALUE
verify-eventenquiry		▬
verify-lodgingenquiry		▬
verify-lodgingreservation		▬
submit-eventenquiry		▬
submit-lodgingenquiry		▬
submit-lodgingreservation		▬
question-activity		▬

Text Categorization

- Take a document and assign it a label representing its content (ACM keyword, Yahoo category)
- Classic example: decide whether a newspaper article is about politics, business, or sports
- There are many other uses for the same technology:
  - Is this page a laser printer product page?
  - Does this company accept overseas orders?
  - What kind of job does this job posting describe?
  - What kind of position does this list of responsibilities describe?
  - What position does this this list of skills best fit?
  - Is this the “computer” or “harbor” sense of *port*?

Text Categorization Problem [Basili&Moschitti'05]

- Given:
  - a set of target categories: C = {$\mathbf{C}^1$, .., $\mathbf{C}^n$}
  - the set *T* of documents,

  define a mapping function *f* from the documents in *T* to the set of categories C

- Vector Space Model
  - Features (e.g. doc words) are dimensions of a Vector Space.
  - Documents and Categories are vectors of feature weights.
  - Vector ***d*** (representing a document) is given class $\mathbf{C}^i$ if

$$\vec{d} \cdot \vec{C}^i > th$$

The Vector Space Model

d_1: Politics

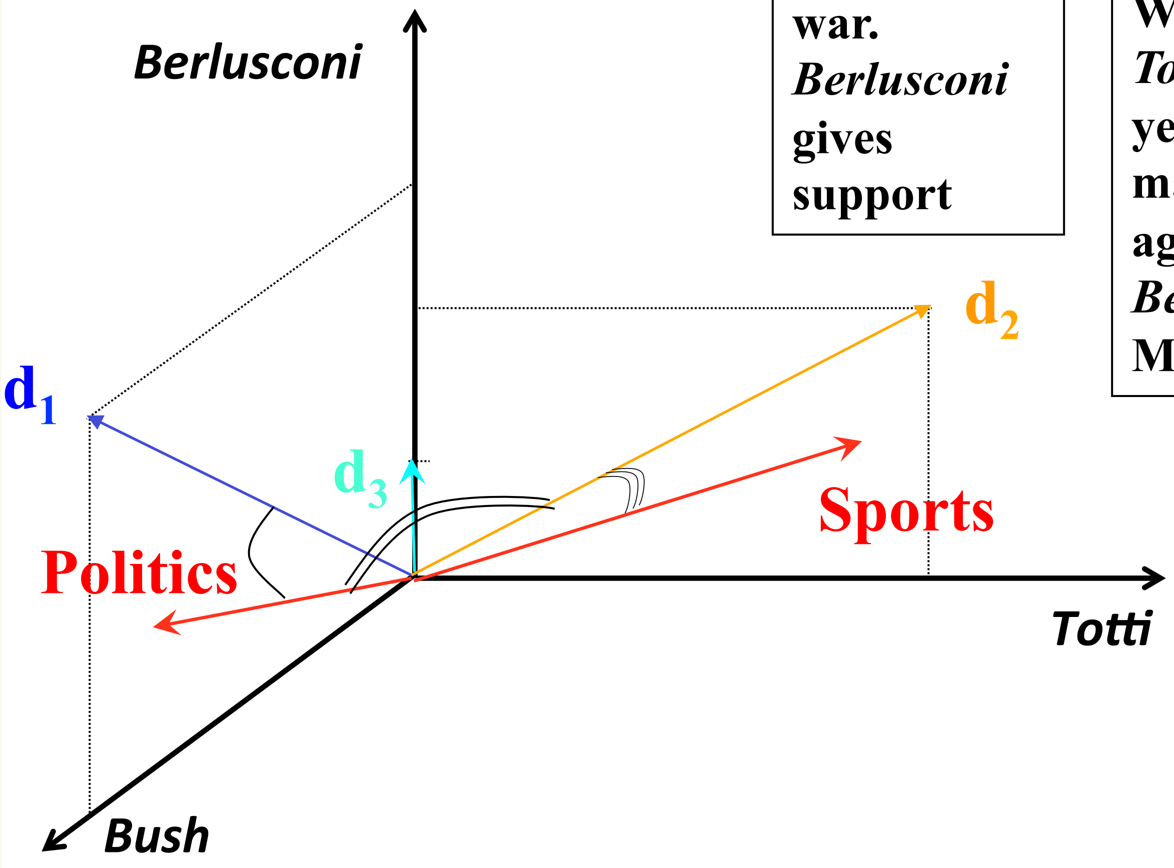
Bush declares war. *Berlusconi* gives support

d_2: Sports

Wonderful *Totti* in yesterday’s match against *Berlusconi’s* Milan

d_3: Economics

Berlusconi acquires Inzaghi before elections



Text Categorization process

- A corpus of pre-categorized documents
- Split document in two parts:
  - Training-set
  - Test-set
- Apply a supervised machine learning model to the training-set
  - Positive examples
  - Negative examples
- Measure the performances on the test-set
  - e.g., Precision and Recall

Feature Vectors

- Each example is associated with a vector of *n* feature types (e.g. unique words in TC)

$$\vec{x} = (0,\ ..,1,..,0,..,0,\ ..,1,..,0,..,0,\ ..,1,..,0,..,0,\ ..,1,..,0,..,\ 1)$$

acquisition buy market sell stocks

- The dot product $\vec{x} \cdot \vec{z}$ counts the number of features in common

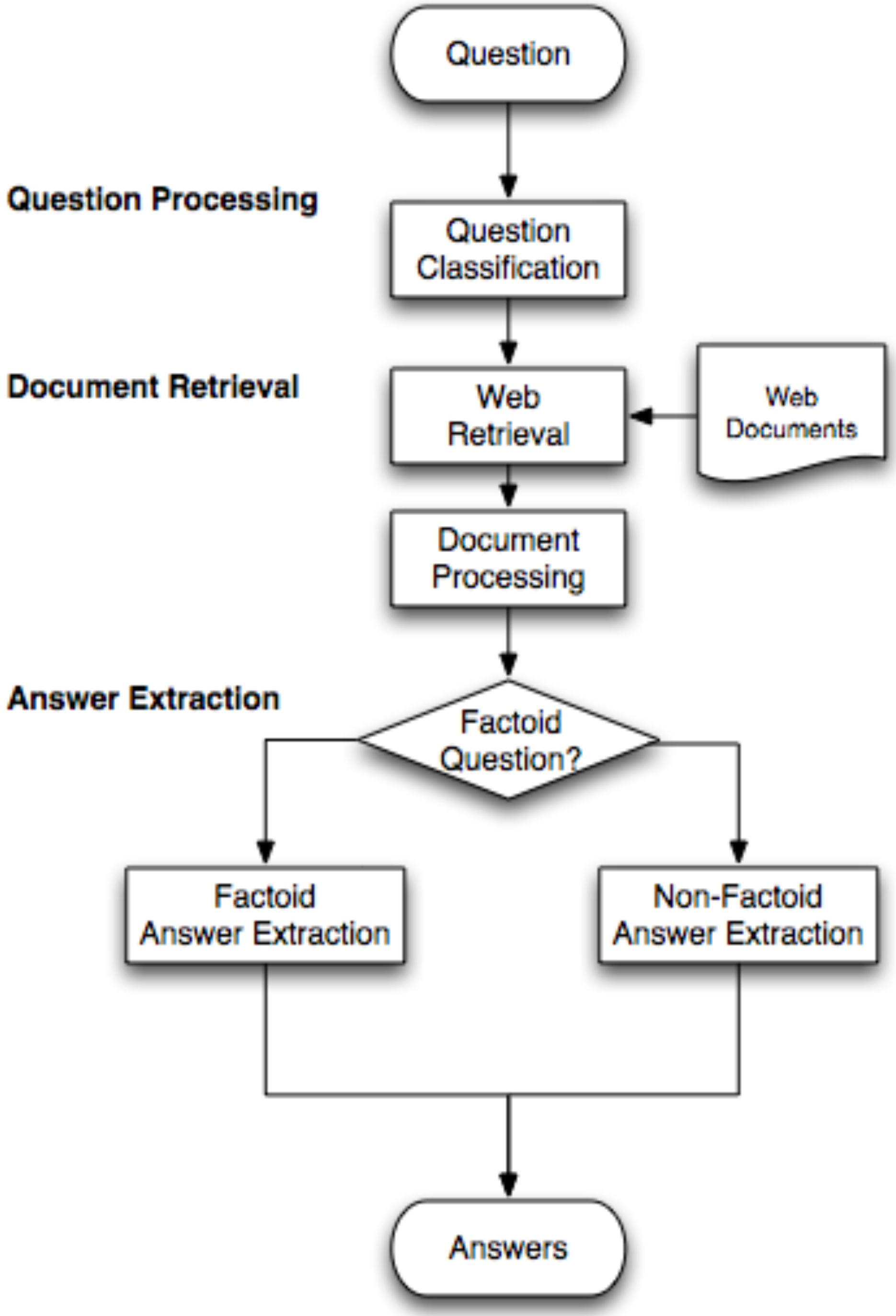
- This provides a sort of *similarity*

Question Answering (QA)

- Automatically answering a natural language question via natural language answers
  - Answer: not documents, but words/phrases/sentences
- [Simmons, 1965]: natural language interfaces to small DBs
- Current QA systems are open-domain
  - Question topic/domain is unconstrained
  - Answer source also unconstrained (e.g. Web)
- Questions come in different types and levels of difficulty
  - Why does the moon turn orange? (non-*factoid*)
  - What is rubella? (non-*factoid*)
  - When was Barack Obama born? (*factoid*)
    - August 4, 1961
  - Who was president when Barack Obama was born?
    - John F. Kennedy
  - How many presidents have there been since Barack Obama was born?
    - 9

YourQA [Quarteroni'08]

- A complex, open-domain question answering system
- Three phases:
  1. Question Classification
  2. Document Retrieval
  3. Answer Extraction

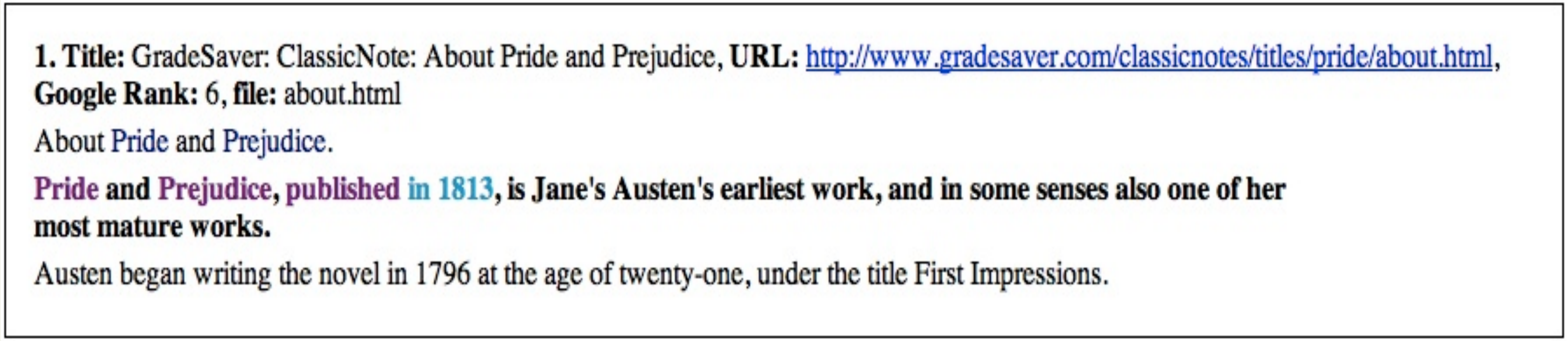


Answer Extraction [NLE'09]

1. Compute match between the query and each retrieved document sentence
   - Weighted sum of lexical/syntactic similarity scores
   - For factoids: NEs/patterns compatible with question class
2. For each document,
   - identify best matching sentence **s***
   - create a passage centred around **s***
3. Rank all passages based on match
4. Return top *m* passages to the user

YourQA result format

“When was *Pride and Prejudice* published?”

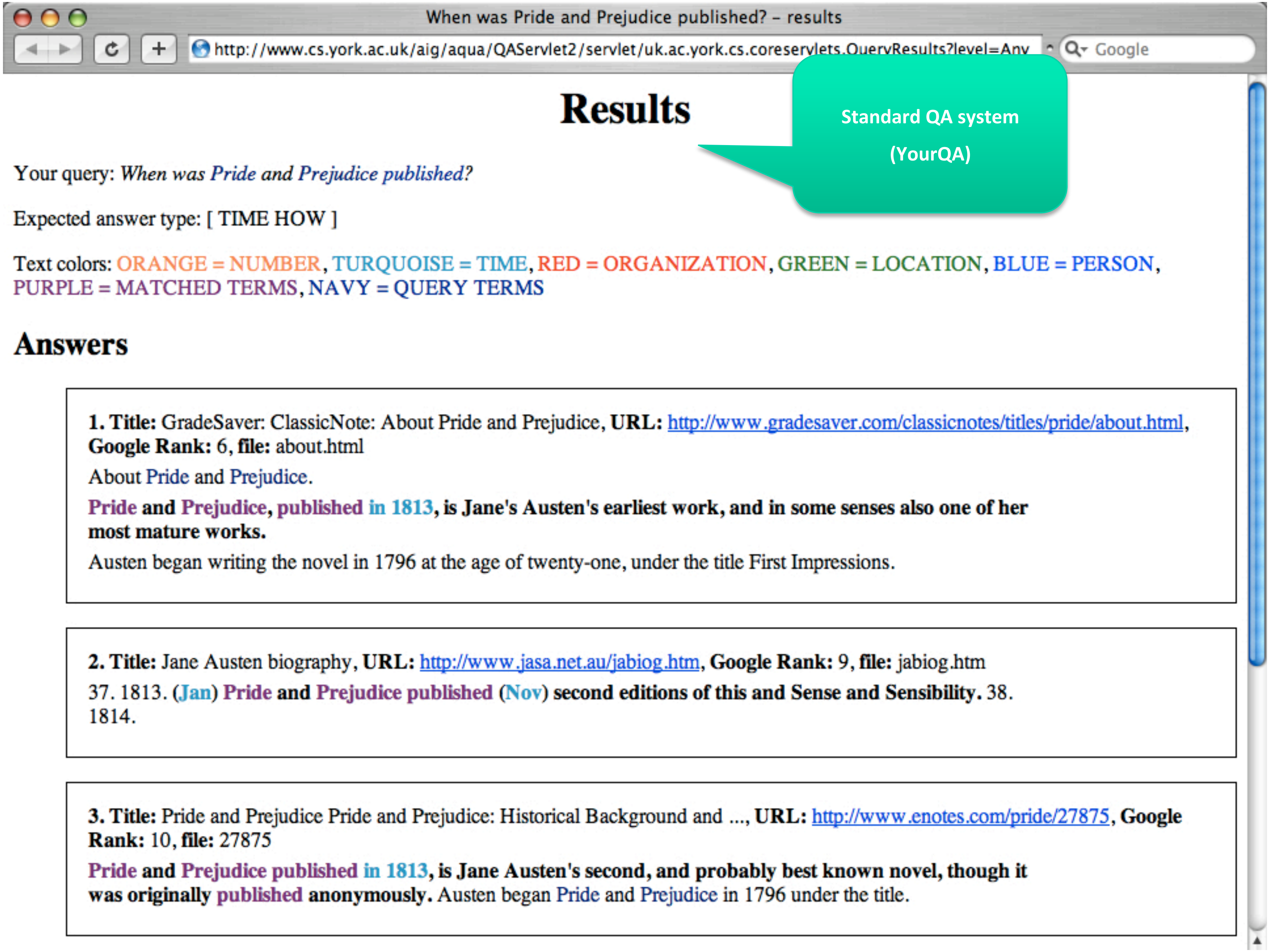


1. Title: GradeSaver: ClassicNote: About Pride and Prejudice, **URL:** http://www.gradesaver.com/classicnotes/titles/pride/about.html, **Google Rank:** 6, **file:** about.html

About Pride and Prejudice.

Pride and Prejudice, published in 1813, is Jane's Austen's earliest work, and in some senses also one of her most mature works.

Austen began writing the novel in 1796 at the age of twenty-one, under the title First Impressions.



When was Pride and Prejudice published? – results

http://www.cs.york.ac.uk/aig/aqua/QAServlet2/servlet/uk.ac.york.cs.coreservlets.QueryResults?level=Any

Standard QA system

(YourQA)

Results

Your query: *When was Pride and Prejudice published?*

Expected answer type: [TIME HOW]

Text colors: ORANGE = NUMBER, TURQUOISE = TIME, RED = ORGANIZATION, GREEN = LOCATION, BLUE = PERSON, PURPLE = MATCHED TERMS, NAVY = QUERY TERMS

Answers

1. Title: GradeSaver: ClassicNote: About Pride and Prejudice, **URL:** http://www.gradesaver.com/classicnotes/titles/pride/about.html, **Google Rank:** 6, **file:** about.html

About Pride and Prejudice.

Pride and Prejudice, published in 1813, is Jane's Austen's earliest work, and in some senses also one of her most mature works.

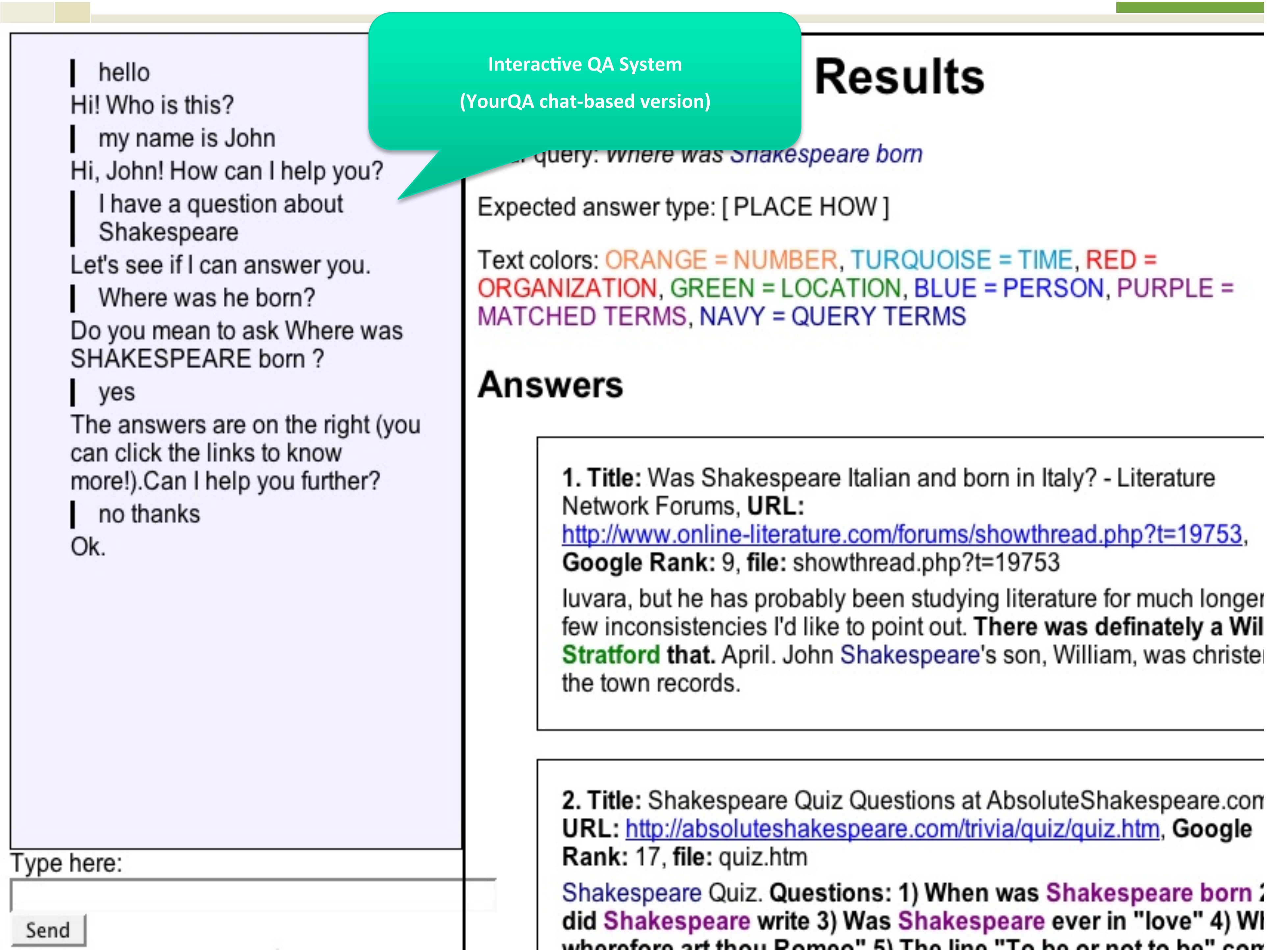
Austen began writing the novel in 1796 at the age of twenty-one, under the title First Impressions.

2. Title: Jane Austen biography, **URL:** http://www.jasa.net.au/jabiog.htm, **Google Rank:** 9, **file:** jabiog.htm

37. 1813. **(Jan) Pride and Prejudice published (Nov) second editions of this and Sense and Sensibility.** 38. 1814.

3. Title: Pride and Prejudice Pride and Prejudice: Historical Background and ..., **URL:** http://www.enotes.com/pride/27875, **Google Rank:** 10, **file:** 27875

Pride and Prejudice published in 1813, is Jane Austen's second, and probably best known novel, though it was originally published anonymously. Austen began Pride and Prejudice in 1796 under the title.



Interactive QA System

(YourQA chat-based version)

hello

Hi! Who is this?

my name is John

Hi, John! How can I help you?

I have a question about Shakespeare

Let's see if I can answer you.

Where was he born?

Do you mean to ask Where was SHAKESPEARE born ?

yes

The answers are on the right (you can click the links to know more!).Can I help you further?

no thanks

Ok.

Results

query: Where was Shakespeare born

Expected answer type: [PLACE HOW]

Text colors: ORANGE = NUMBER, TURQUOISE = TIME, RED = ORGANIZATION, GREEN = LOCATION, BLUE = PERSON, PURPLE = MATCHED TERMS, NAVY = QUERY TERMS

Answers

1. Title: Was Shakespeare Italian and born in Italy? - Literature Network Forums, **URL:** http://www.online-literature.com/forums/showthread.php?t=19753, **Google Rank:** 9, **file:** showthread.php?t=19753

Iuvara, but he has probably been studying literature for much longe
few inconsistencies I'd like to point out. **There was definately a Wi**
Stratford that. April. John Shakespeare's son, William, was christe
the town records.

2. Title: Shakespeare Quiz Questions at AbsoluteShakespeare.co
URL: http://absoluteshakespeare.com/trivia/quiz/quiz.htm, **Google Rank:** 17, **file:** quiz.htm

Shakespeare Quiz. **Questions: 1) When was Shakespeare born**
did Shakespeare write 3) Was Shakespeare ever in "love" 4) W

Complex Question Answering

- Answering some types of questions is challenging
  - Complex semantics, short texts, data sparseness
- Example: definition questions
  - “*What are* ***antigens****?*”, “*What is* ***autism****?*”
- Word match with query is not enough to assess whether candidate answer is a correct definition
  - **OK*:*** ***Autism*** *is a behavioral disorder*
  - **Not OK*:*** ***Autism*** *affects thousands of Americans*
- Other useful linguistic features are *structured:*
  - Syntactic parse trees
  - Shallow semantic trees

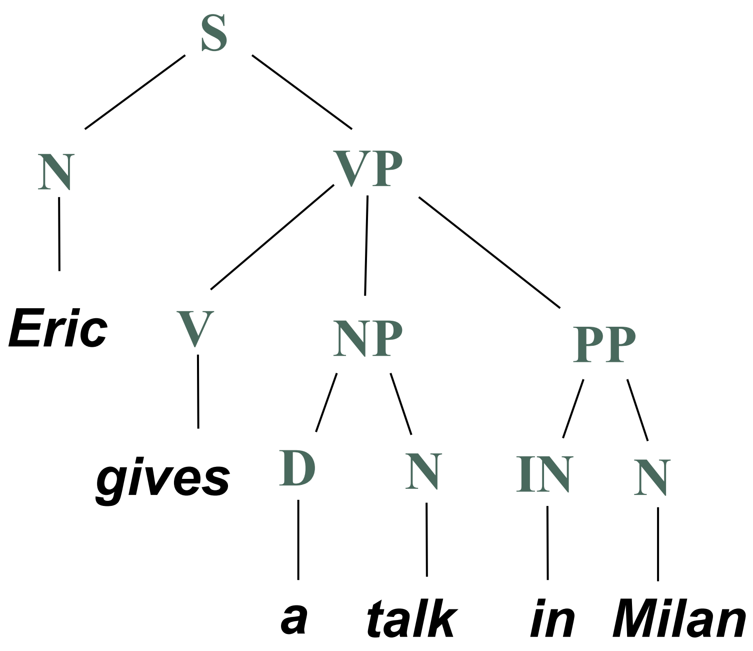
Complex Question Answering
[IPM'10]

- Further to "classic" QA algorithm, answer classification & re-ranking may be applied
- In this presentation:
  - Machine learning methods (SVMs) to train complex (e.g. definition) answer classifiers
  - Structural text representations used as features
    - Syntactic parse trees
    - Shallow semantic parse trees

Syntactic parse trees (PT)

- A well-known source of syntactic information

"Eric gives a talk in Milan"



Shallow Semantic parsing:
Predicate Argument Structures (PAS)

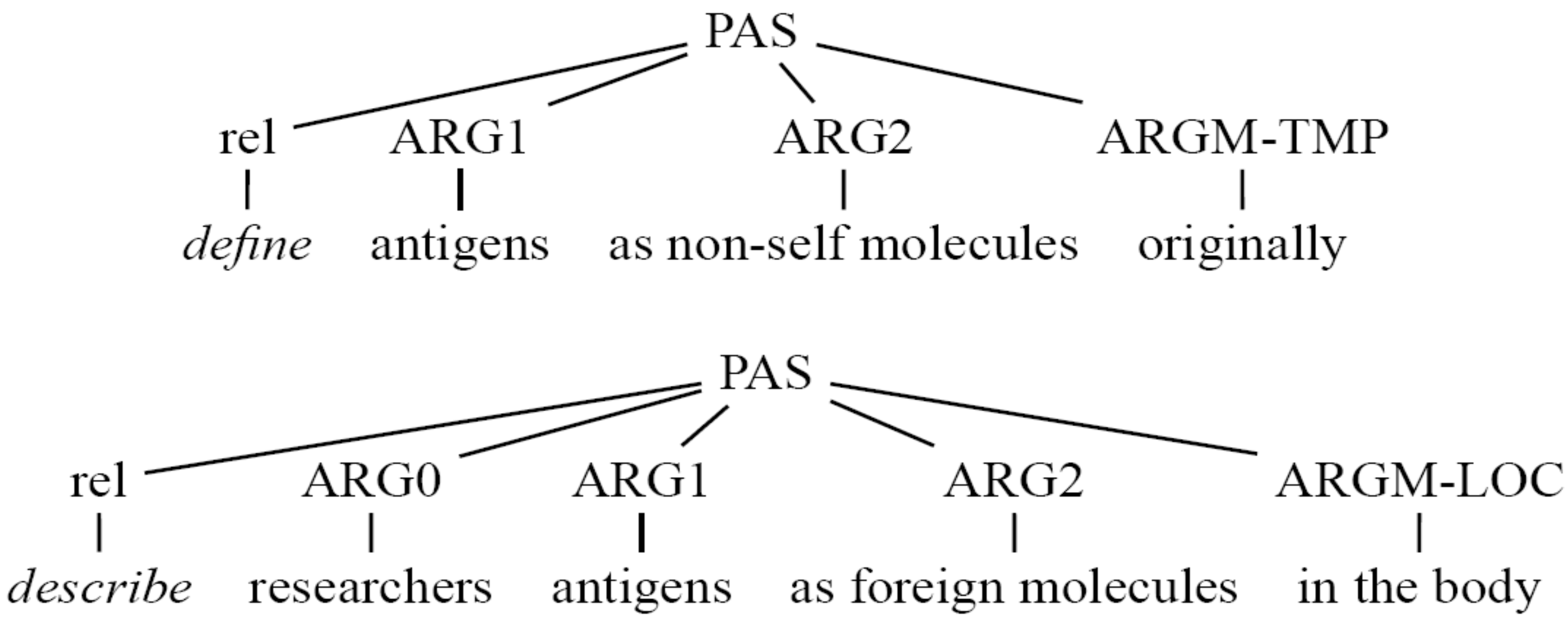
- Given an event:
  - predicates describe a relation among entities
  - entities are called arguments

- Examples:
  1. *Antigens were originally defined as non-self molecules.*
  2. *Researchers describe antigens as foreign molecules in the body.*

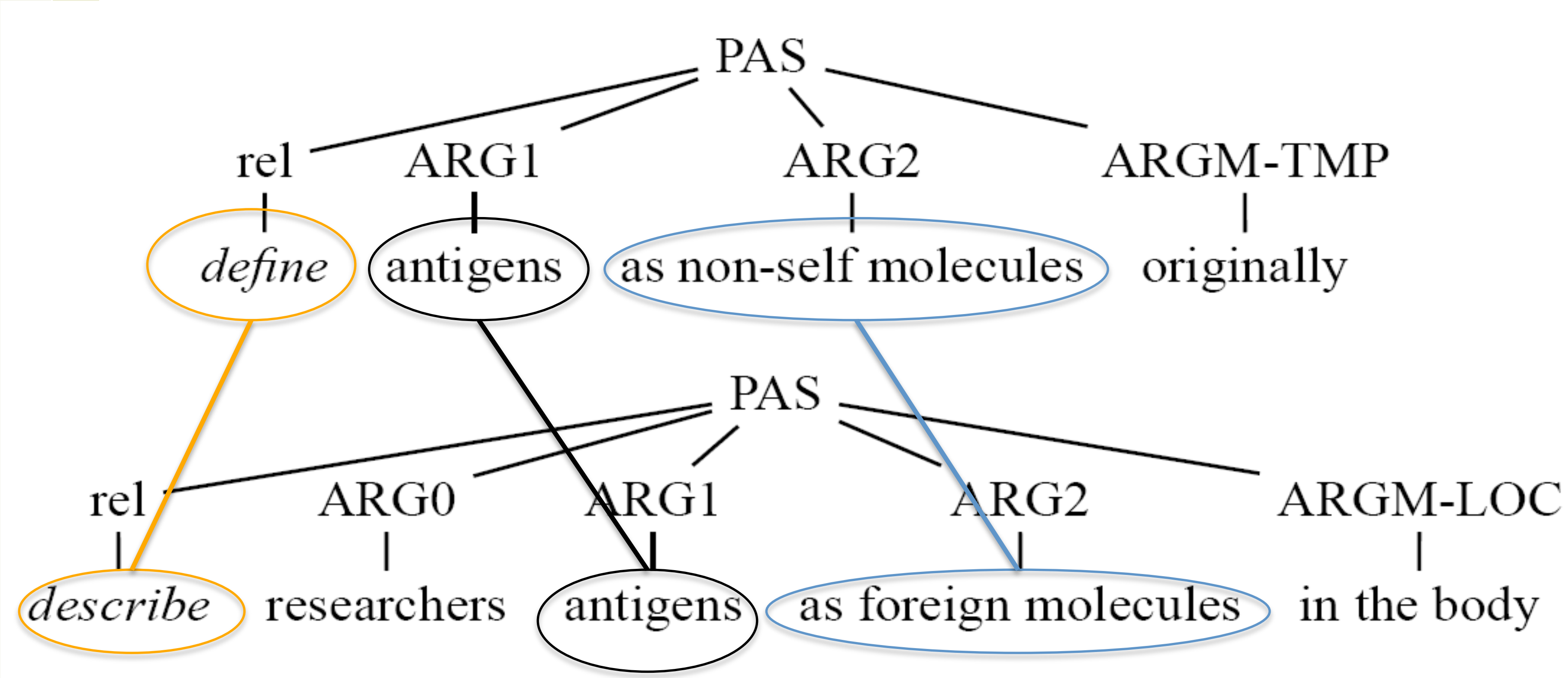
Shallow Semantic parsing:
Predicate Argument Structures (PAS)

- Given an event:
  - predicates describe a relation among entities
  - entities are called arguments
- Examples:
  1. **[A1** *Antigens***]** *were* **[A−TMP** *originally***]** **[*rel*** *defined***]** **[A2** *as non-self molecules***]**.
  2. **[A0** *Researchers***]** **[*rel*** *describe***]** **[A1** *antigens***]** **[A2** *as foreign molecules***]** .

Tree Representation of a PAS



Tree Representation of a PAS



Tree Kernel Methods

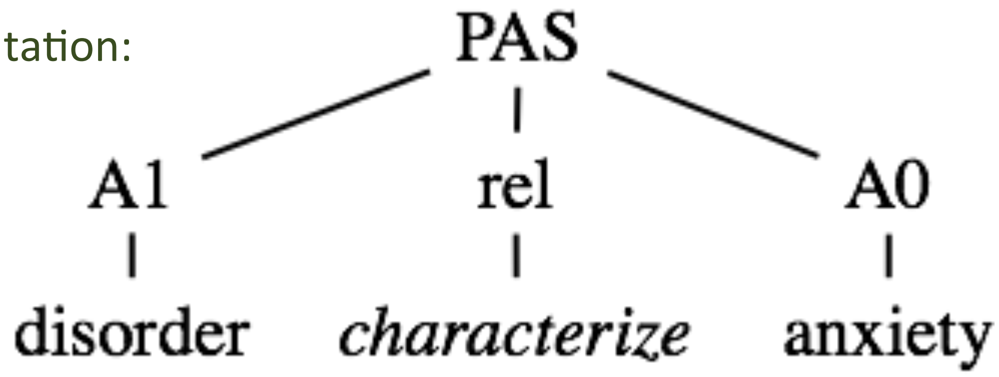
- Idea: given the tree representations of 2 sentences, we can enumerate their common substructures using various kernel functions.
- We can learn to label new instances by comparing them with instances from our training corpus on the grounds of common syntactic/semantic substructures.

Explicit feature spaces created by Kernel functions

TREC question: “What is autism?”

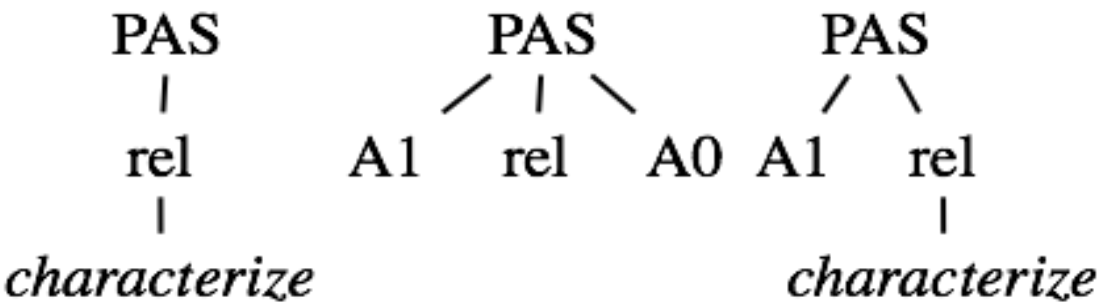
Candidate answer from YourQA: “Autism is a disorder characterized by social anxiety”

Results in the PAS tree representation:



Partial Tree Kernel function enumerates PAS tree substructures for similarity computation.

Examples:



Answer Classification Experiments

[IPM'10]

- Task: binary classification
  - Instance: ***<question, candidate answer>*** pairs
  - Questions: 138 TREC 2001 "description" questions
- Two answer datasets:
  - WEB: from the Web, 1309 sentences, 416 judged correct
  - TREC: news (AQUAINT-6 corpus), 2256 sentences, 261 correct
- Learning models: SVMs, various kernel functions
- Data representations: bag of words, syntax trees, POS sequences, Predicate-argument structures

Interesting findings on PAS tree kernels

- Powerful in finding definition patterns such as
  - [**A1** X] [**R-A1** *that*] [***rel*** *result*] [**A2** Y]
  - [**A1** X] [***rel*** *characterize*] [**A0** Z]
- Able to generalize utterances such as
  - *"**German measles**, that result in red marks on the skin, are a common disease in children."*
  - *"**Autism** is characterized by the inability to relate to other people."*

Answer Re-ranking [ACL'07]

- Output of binary classifier re-orders top answers
  - negative SVM prediction => lower ranking
- Mean Reciprocal Rank (MRR) on Web-QA dataset:

	Google	YourQA	Re-ranker
MRR@5	48.9 ± 3.8	56.2 ± 3.2	81.1 ± 2.1

IBM Watson destroys all humans at *Jeopardy!*

- *Jeopardy!* is a legendary US TV game
- Given a hint, guess question
  - Hint: This number, one of the first 20, uses only one vowel (4 times!).
    - Solution: Seventeen
  - Hint: Sakura cheese from Hokkaido is a soft cheese flavored with leaves from this fruit tree.
    - Solution: Cherry
- It's awfully hard to play this game for a computer
  - Clues are often ambiguous/wordplay
  - It's difficult to classify clues based on expected response
- But IBM TJ Watson DeepQA lab designed a challenger, **Watson**, that effectively beat the best two human contestants ever!
  - IBM experts blended probabilistic & linguistic features using a Question Answering approach

Conclusions

- Complete human-level natural language understanding is still a distant goal
- But there are now practical and usable partial NLU systems applicable to many problems
- An important design decision is in finding an appropriate match between (parts of) the application domain and the available methods
- *But, used with care, statistical NLP methods have opened up new possibilities for high performance text understanding systems.*

References

- [Wiezenbaum, 1966] J. Wiezenbaum. ELIZA - A Computer Program for the Study of Natural Language Communication between Man and Machine, *Communications of the ACM* 9 (1966): 36-45.
- [Winograd, 1971] T. Winograd. Procedures as a Representation for Data in a Computer Program for Understanding Natural Language. MIT AI Technical Report 235, February 1971
- [Schank & Abelson, 1977] Schank, Roger C., and Robert P. Abelson. 1977. Scripts, plans, goals, and understanding: An inquiry into human knowledge structures. Hillsdale, NJ: Lawrence Erlbaum.
- [Joachims, 1998] Text categorization with Support Vector Machines: Learning with many relevant features.Machine Learning: ECML-98. Springer LNCS vol 1398/1998, pp. 137-142
- [Gorin et al.,1997] A.L Gorin, G Riccardi, J.H Wright, How may I help you?, Speech Communication, 23(1-2) 1997, pp.113-127
- [Finkel et al.,2005] J. R. Finkel, T. Grenager, C. Manning. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. Proceedings of ACL 2005, pp. 363-370.
- [Carreras & Marquez, 2005] X. Carreras, L. Marquez. Introduction to the CoNLL-2005 shared task: semantic role labeling. Proc. CoNLL shared task, ACL, 2005.
- [Schmid, 1994] H. Schmid, Probabilistic Part-of-Speech Tagging using Decision Trees. In Proceedings of the International Conference on New Methods in Language Processing (1994), pp. 44-49.

References (2)

- [Basili & Moschitti'05] R. Basili and A. Moschitti (2005), Automatic Text Categorization: from Information Retrieval to Support Vector Learning. Aracne Editrice, Roma.
- [Turney'02] P. Turney (2002). *Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews*. Proc. ACL'2002.
- [Wong et al.'09] W. Wong and W. Liu and M. Bennamoun (2009). Acquiring Semantic Relations using the Web for Constructing Lightweight Ontologies. *Proc. PAKDD*.
- [Zelenko et al.'03] D. Zelenko, C. Aone and A. Richardella (2003). Kernel Methods for Relation Extraction. Journal of Machine Learning Research 3 (2003) 1083-1106.
- [Giunchiglia et al.,2010] Giunchiglia, F. and Maltese, V. and Farazi, F. and Dutta, B. GeoWordNet: a resource for geo-spatial applications*.* Proc. ESWC 2010.

References (3)

- [Quarteroni’08] S. Quarteroni, Advanced Techniques for Personalized, Interactive Question Answering. PhD thesis, University of York, UK, 2008
- [NLE’09] S. Quarteroni and S. Manandhar, Designing an Interactive Open Domain Question Answering System. In: Natural Language Engineering, Vol. 15(1), pp. 73-95. N. Webb, B. Webber, eds., Cambridge University Press, 2009.
- [IPM’10] A. Moschitti and S. Quarteroni, Linguistic Kernels for Answer Re-ranking in Question Answering Systems. Information Processing & Management, Special Issue on Question Answering, 2010. In press.
- [ACL’07] A. Moschitti, S. Quarteroni, R. Basili and S. Manandhar, Exploiting Syntactic and Shallow Semantic Kernels for Question/Answer Classification. In: Proceedings of ACL’07, Prague, Czech Republic, June 2007.
- [ACL’08] A. Moschitti, S. Quarteroni, Kernels on Linguistic Structures for Answer Extraction, Short paper in: Proceedings of ACL'08, Columbus, OH, USA, June 2008.