Information Extraction and Named Entity Recognition

Introducing the tasks:

Getting simple structured information out 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:
    - *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)

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

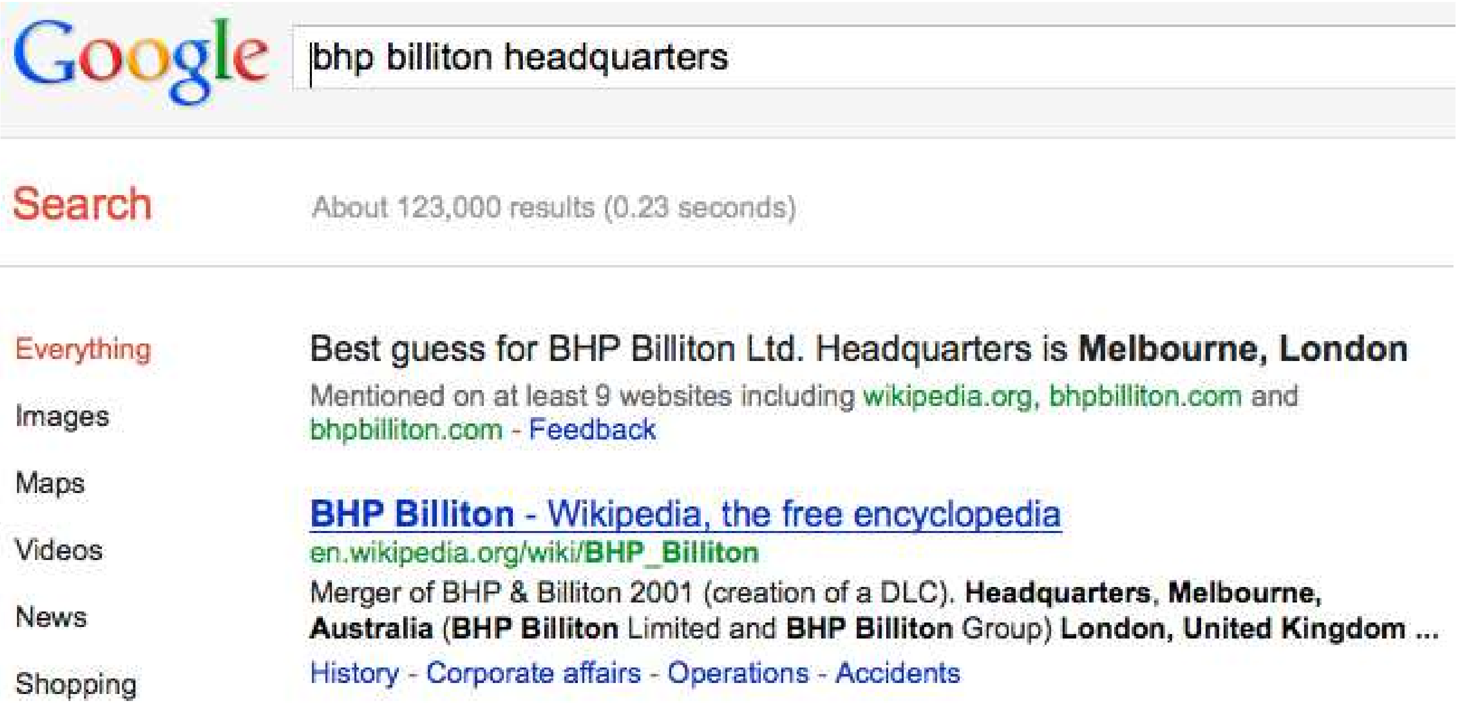
Low-level information extraction

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

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

Christopher Manning

Low-level information extraction



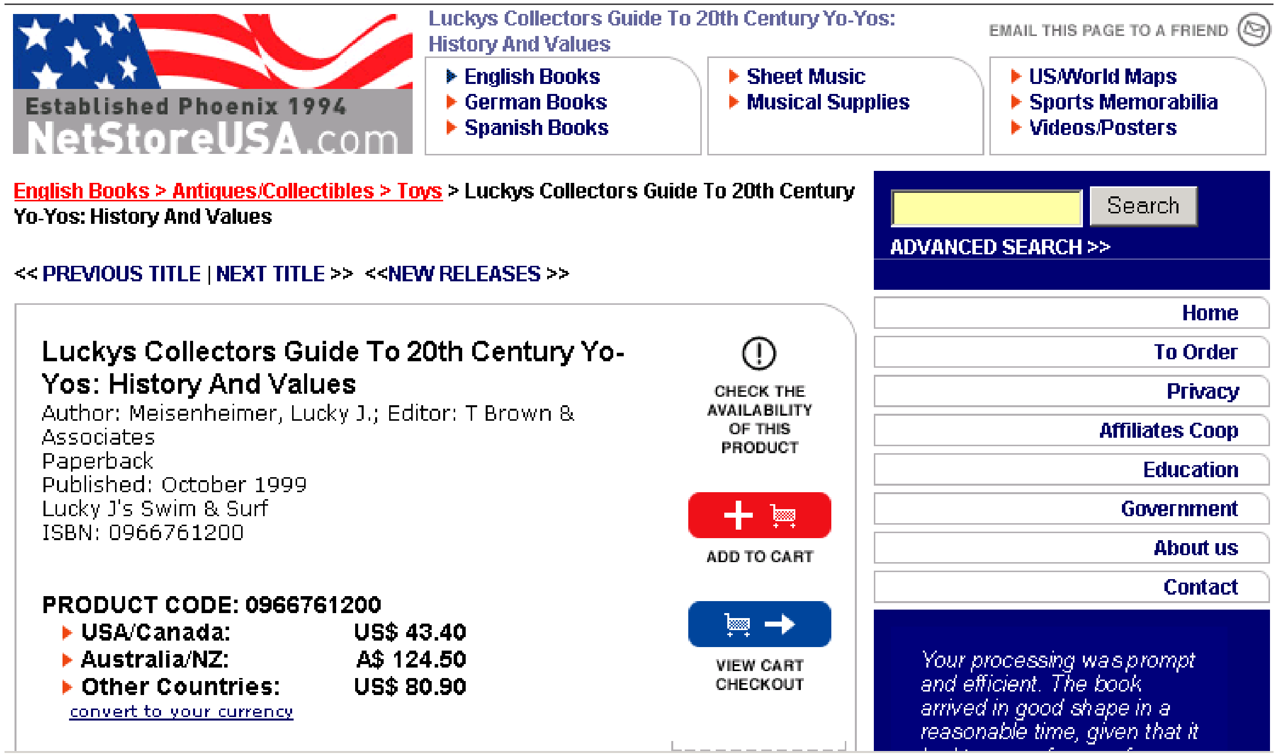
Christopher Manning

Why is IE hard on the web?

A book,
Not a toy

Title

Need this
price



How is IE useful?

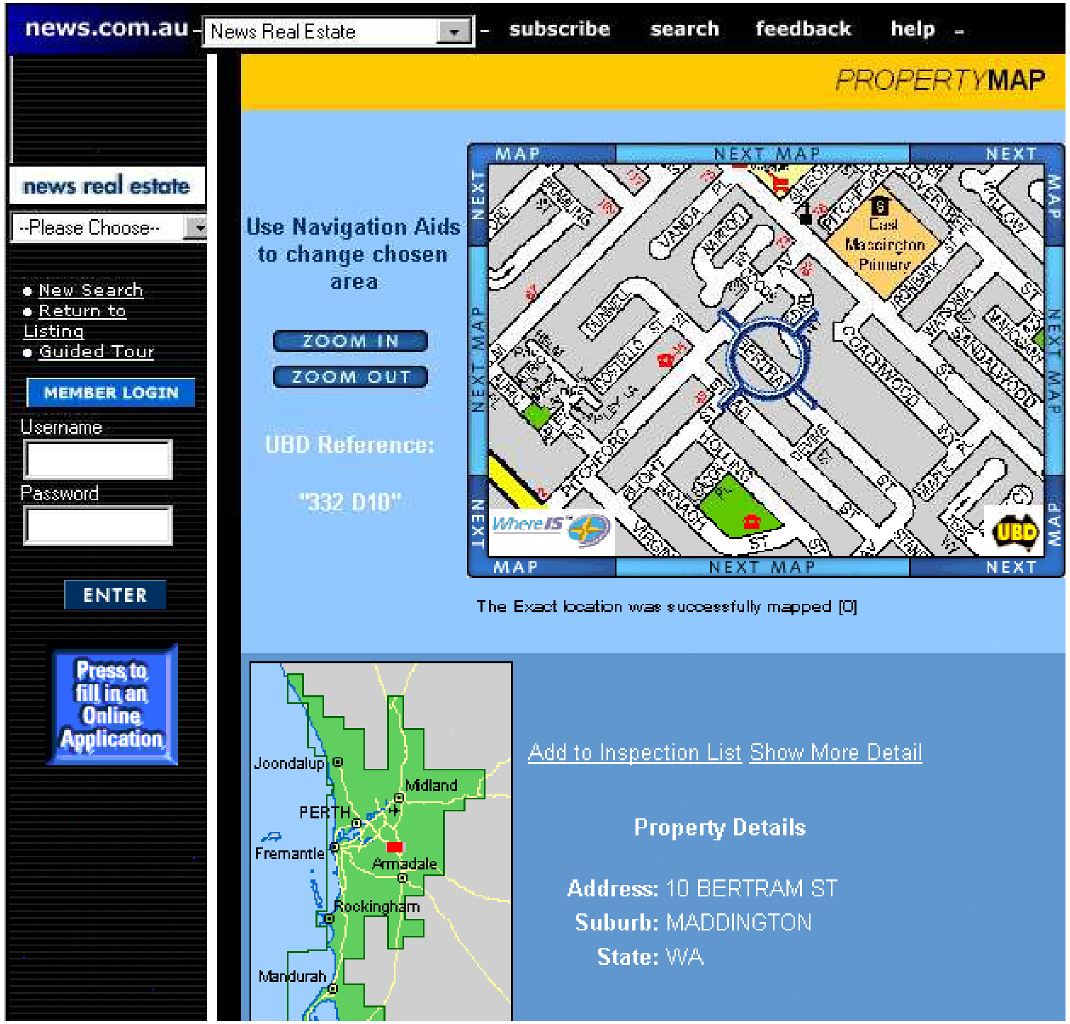
Classified Advertisements (Real Estate)

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>
```

Christopher Manning



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 Parramatta
  - 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

Named Entity Recognition (NER)

- A very important sub-task: find and classify names in text, for example:
  - The decision by the independent MP Andrew Wilkie to withdraw his support for the minority Labor government sounded dramatic but it should not further threaten its stability. When, after the 2010 election, Wilkie, Rob Oakeshott, Tony Windsor and the Greens agreed to support Labor, they gave just two guarantees: confidence and supply.

Named Entity Recognition (NER)

- A very important sub-task: find and classify names in text, for example:

  - The decision by the independent MP Andrew Wilkie to withdraw his support for the minority Labor government sounded dramatic but it should not further threaten its stability. When, after the 2010 election, Wilkie, Rob Oakeshott, Tony Windsor and the Greens agreed to support Labor, they gave just two guarantees: confidence and supply.

Named Entity Recognition (NER)

- A very important sub-task: find and classify names in text, for example:
    - The decision by the independent MP Andrew Wilkie to withdraw his support for the minority Labor government sounded dramatic but it should not further threaten its stability. When, after the 2010 election, Wilkie, Rob Oakeshott, Tony Windsor and the Greens agreed to support Labor, they gave just two guarantees: confidence and supply.

Person
Date
Location
Organi-
zation

Christopher Manning

Named Entity Recognition (NER)

- The uses:
  - Named entities can be indexed, linked off, etc.
  - Sentiment can be attributed to companies or products
  - A lot of IE relations are associations between named entities
  - For question answering, answers are often named entities.
- Concretely:
  - Many web pages tag various entities, with links to bio or topic pages, etc.
    - Reuters' OpenCalais, Evri, AlchemyAPI, Yahoo's Term Extraction, …
  - Apple/Google/Microsoft/… smart recognizers for document content

Information Extraction and Named Entity Recognition

Introducing the tasks:

Getting simple structured information out of text

Evaluation of Named Entity Recognition

Precision, Recall, and the
F measure;

their extension to sequences

The 2-by-2 contingency table

	correct	not correct
selected	tp	fp
not selected	fn	tn

Precision and recall

- **Precision**: % of selected items that are correct
  Recall: % of correct items that are selected

	correct	not correct
selected	tp	fp
not selected	fn	tn

A combined measure: F

- A combined measure that assesses the P/R tradeoff is F measure (weighted harmonic mean):

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

- The harmonic mean is a very conservative average; see *IIR* § 8.3
- People usually use balanced F1 measure
  - i.e., with $\beta = 1$ (that is, $\alpha = ½$): $F = 2PR/(P+R)$

Quiz question

What is the F_1?

$P = 40\%$ $R = 40\%$ $F_1 =$

Quiz question

What is the F_1?

$P = 75\%$ $R = 25\%$ $F_1 =$

Christopher Manning

The Named Entity Recognition Task

Task: Predict entities in a text

Foreign	ORG
Ministry	ORG
spokesman	O
Shen	PER
Guofang	PER
told	O
Reuters	ORG
:	:

Standard evaluation is per entity, *not* per token

Precision/Recall/F1 for IE/NER

- Recall and precision are straightforward for tasks like IR and text categorization, where there is only one grain size (documents)
- The measure behaves a bit funnily for IE/NER when there are *boundary errors* (which are *common*):
  - First Bank of Chicago announced earnings …
- This counts as both a fp and a fn
- Selecting *nothing* would have been better
- Some other metrics (e.g., MUC scorer) give partial credit (according to complex rules)

Evaluation of Named Entity Recognition

Precision, Recall, and the F measure;

their extension to sequences

Methods for doing NER and IE

The space;

hand-written patterns

Three standard approaches to NER (and IE)

1. Hand-written regular expressions
   - Perhaps stacked
2. Using classifiers
   - Generative: Naïve Bayes
   - Discriminative: Maxent models
3. Sequence models
   - HMMs
   - CMMs/MEMMs
   - CRFs

Hand-written Patterns for Information Extraction

- If extracting from automatically generated web pages, simple regex patterns usually work.
  - Amazon page
  - <div class="buying"><h1 class="parseasinTitle">(.*?)</h1>
- For certain restricted, common types of entities in unstructured text, simple regex patterns also usually work.
  - Finding (US) phone numbers
  - (?:\(?[0-9]{3}\)?[-.])?[0-9]{3}[-.]?[0-9]{4}

Christopher Manning

MUC: the NLP genesis of IE

- DARPA funded significant efforts in IE in the early to mid 1990s
- Message Understanding Conference (MUC) was an annual event/competition where results were presented.
- Focused on extracting information from news articles:
  - Terrorist events
  - Industrial joint ventures
  - Company management changes
- Starting off, all rule-based, gradually moved to ML

MUC Information Extraction Example

Bridgestone Sports Co. said Friday it had set up a joint venture in Taiwan with a local concern and a Japanese trading house to produce golf clubs to be supplied to Japan. The joint venture, Bridgestone Sports Taiwan Co., capitalized at 20 million new Taiwan dollars, will start production in January 1990 with production of 20,000 iron and "metal wood" clubs a month.

JOINT-VENTURE-1

- Relationship: TIE-UP
- Entities: "Bridgestone Sport Co." , "a local concern", "a Japanese trading house"
- Joint Ent: "Bridgestone Sports Taiwan Co."
- Activity: ACTIVITY-1
- Amount: NT$20 000 000

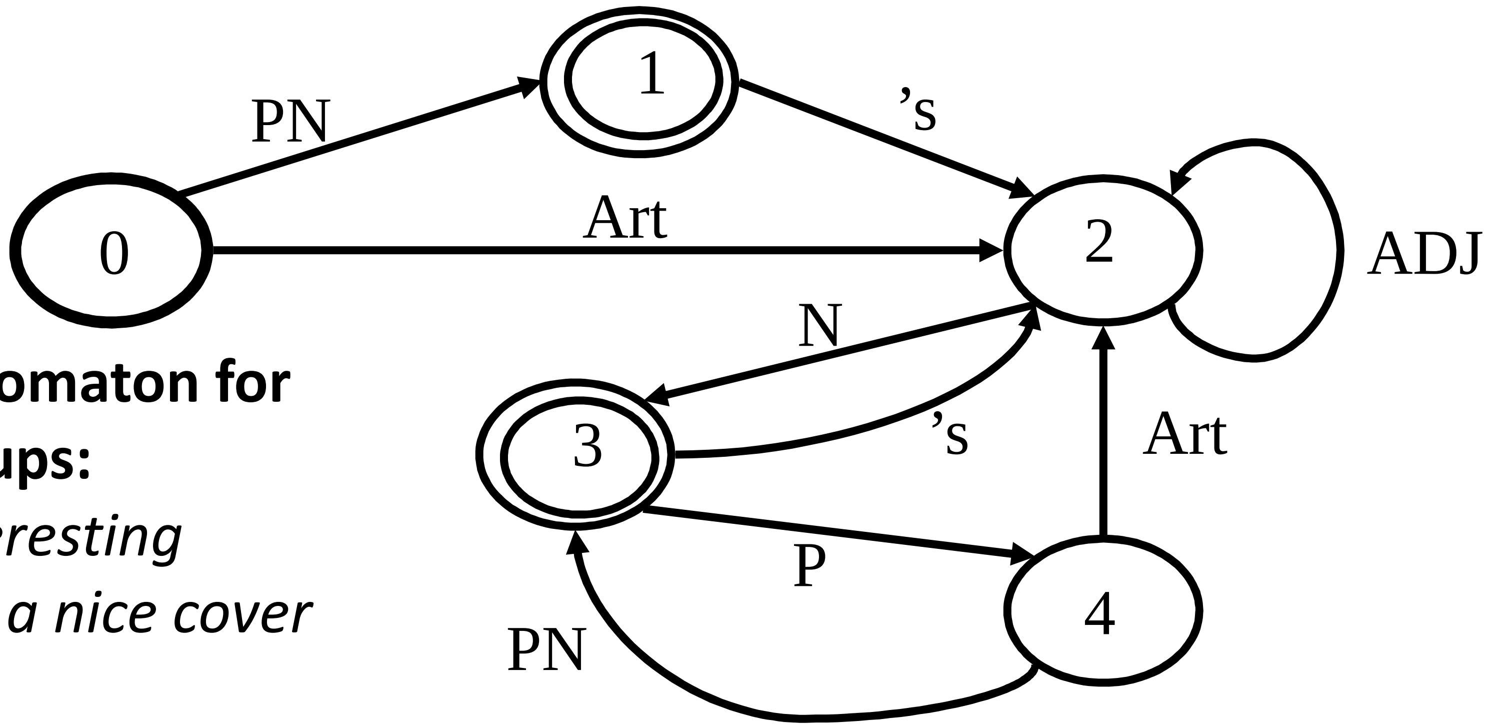
ACTIVITY-1

- Activity: PRODUCTION
- Company: "Bridgestone Sports Taiwan Co."
- Product: "iron and 'metal wood' clubs"
- Start date: DURING: January 1990

Natural Language Processing-based Hand-written Information Extraction

- For unstructured human-written text, some NLP may help
  - Part-of-speech (POS) tagging
    - Mark each word as a noun, verb, preposition, etc.
  - Syntactic parsing
    - Identify phrases: NP, VP, PP
  - Semantic word categories (e.g. from WordNet)
    - KILL: kill, murder, assassinate, strangle, suffocate

Grep++ = Cascaded grepping



Finite Automaton for Noun groups:

John's interesting book with a nice cover

Natural Language Processing-based Hand-written Information Extraction

- We use a cascaded regular expressions to match relations
  - Higher-level regular expressions can use categories matched by lower-level expressions
  - E.g. the CRIME-VICTIM pattern can use things matching NOUN-GROUP
- This was the basis of the SRI FASTUS system in later MUCs
- Example extraction pattern
  - Crime victim:
    - Prefiller: [POS: V, Hypernym: KILL]
    - Filler: [Phrase: NOUN-GROUP]

Rule-based Extraction Examples

Determining which person holds what office in what organization

- [person] , [office] *of* [org]
  - Vuk Draskovic, leader of the Serbian Renewal Movement
- [org] (*named*, *appointed*, etc.) [person] Prep [office]
  - NATO appointed Wesley Clark as Commander in Chief

Determining where an organization is located

- [org] *in* [loc]
  - NATO headquarters in Brussels
- [org] [loc] (*division*, *branch*, *headquarters*, etc.)
  - KFOR Kosovo headquarters

Methods for doing NER and IE

The space;

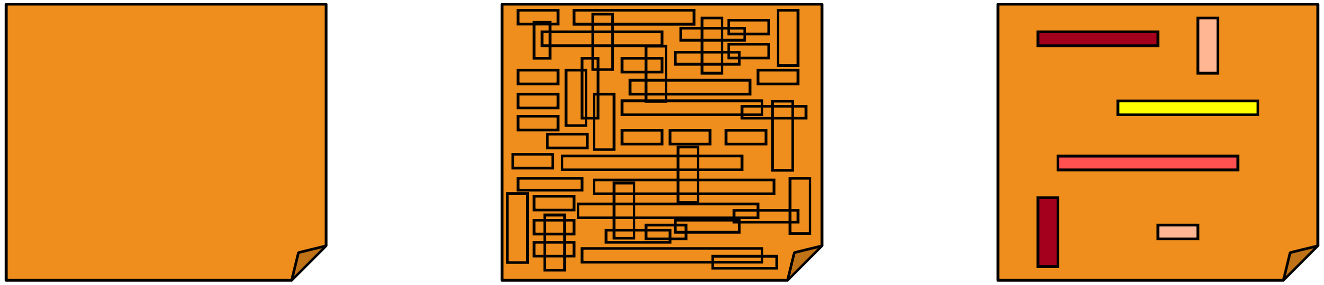
hand-written patterns

Information extraction as text classification

Christopher Manning

Naïve use of text classification for IE

- Use conventional classification algorithms to classify substrings of document as "*to be extracted*" or not.



- In some simple but compelling domains, this naive technique is remarkably effective.
  - But do think about when it would and wouldn't work!

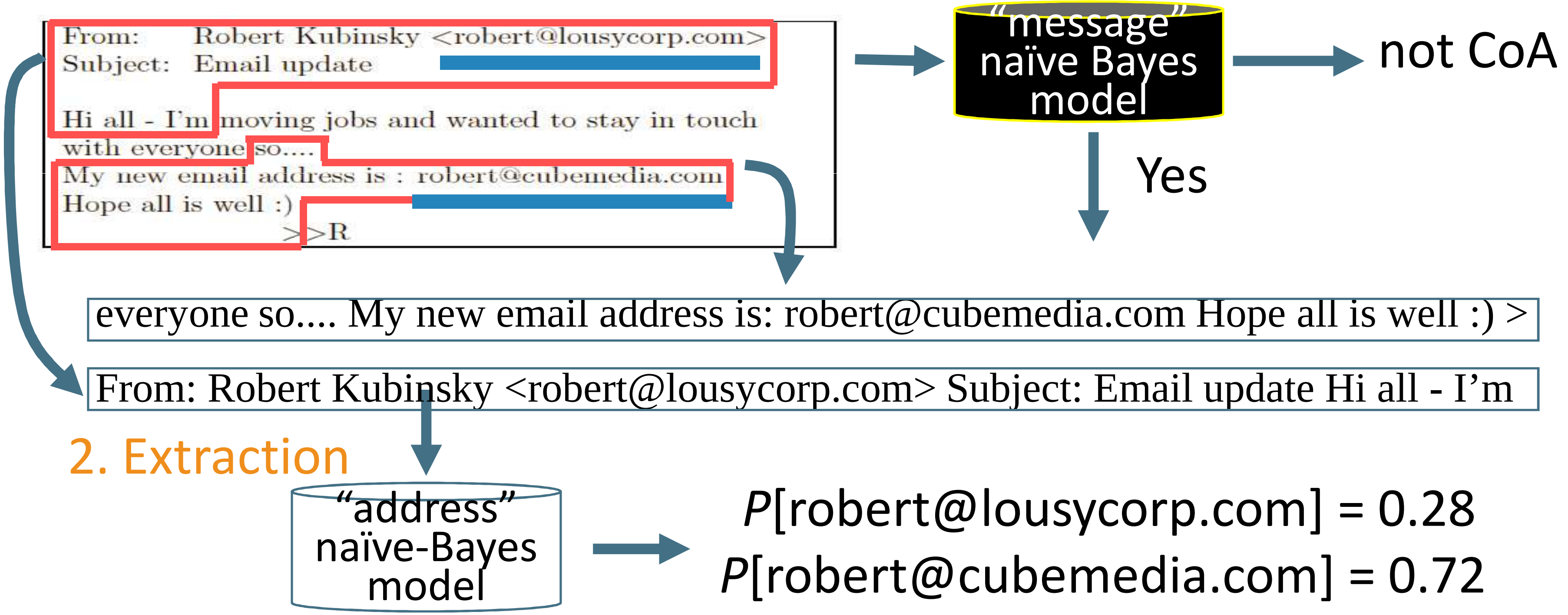
‘Change of Address’ email

From: Robert Kubinsky <robert@lousycorp.com>
Subject: Email update

Hi all - I’m moving jobs and wanted to stay in touch
with everyone so....
My new email address is : $\boxed{\text{robert@cubemedia.com}}$
Hope all is well :)
>>R

Change-of-Address detection
[Kushmerick et al., ATEM 2001]

1. Classification



Change-of-Address detection results [Kushmerick et al., ATEM 2001]

- Corpus of 36 CoA emails and 5720 non-CoA emails
  - Results from 2-fold cross validations (train on half, test on other half)
  - Very skewed distribution intended to be realistic
  - Note very limited training data: only 18 training CoA messages per fold
  - 36 CoA messages have 86 email addresses; old, new, and miscellaneous

	P	R	F_1
Message classification	98%	97%	98%
Address classification	98%	68%	80%

Information extraction as text classification

Sequence Models for Named Entity Recognition

The ML sequence model approach to NER

Training

1. Collect a set of representative training documents
2. Label each token for its entity class or other (O)
3. Design feature extractors appropriate to the text and classes
4. Train a sequence classifier to predict the labels from the data

Testing

1. Receive a set of testing documents
2. Run sequence model inference to label each token
3. Appropriately output the recognized entities

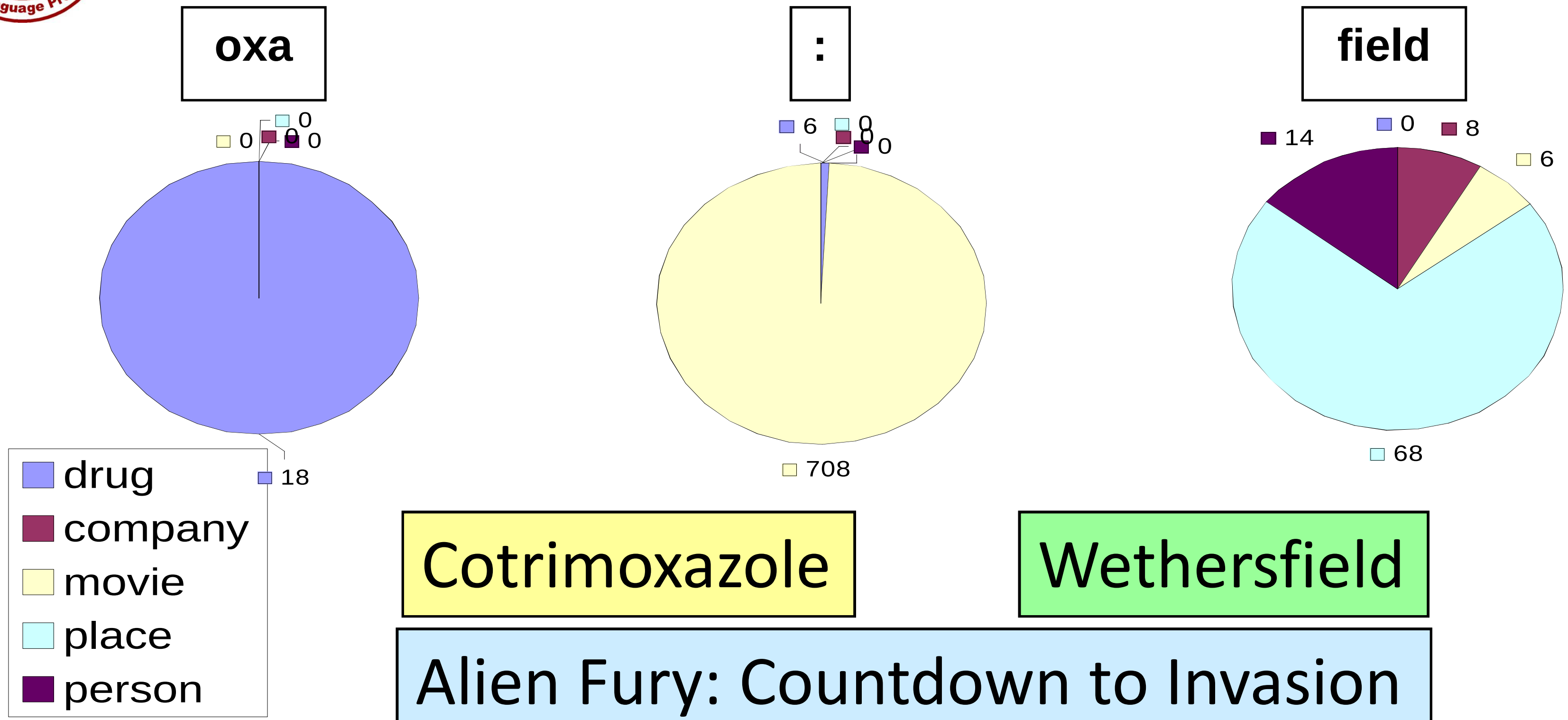
Encoding classes for sequence labeling

	IO encoding	IOB encoding
Fred	PER	B-PER
showed	O	O
Sue	PER	B-PER
Mengqiu	PER	B-PER
Huang	PER	I-PER
‘s	O	O
new	O	O
painting	O	O

Features for sequence labeling

- Words
  - Current word (essentially like a learned dictionary)
  - Previous/next word (context)
- Other kinds of inferred linguistic classification
  - Part-of-speech tags
- Label context
  - Previous (and perhaps next) label

Features: Word substrings



Cotrimoxazole

Wethersfield

Alien Fury: Countdown to Invasion

Features: Word shapes

- Word Shapes
  - Map words to simplified representation that encodes attributes such as length, capitalization, numerals, Greek letters, internal punctuation, etc.

Varicella-zoster	Xx-xxx
mRNA	xXXX
CPA1	XXXd

Sequence Models for Named Entity Recognition

Maximum entropy sequence models

Maximum entropy Markov models (MEMMs) or Conditional Markov models

Sequence problems

- Many problems in NLP have data which is a sequence of characters, words, phrases, lines, or sentences …
- We can think of our task as one of labeling each item

VBG	NN	IN	DT	NN	IN	NN
Chasing	opportunity	in	an	age	of	upheaval

POS tagging

B	B	I	I	B	I	B	I	B	B
而	相	对	于	这	些	品	牌	的	价

Word segmentation

PERS	O	O	O	ORG	ORG
Murdoch	discusses	future	of	News	Corp.

Named entity recognition

Text segmentation

MEMM inference in systems

- For a Conditional Markov Model (CMM) a.k.a. a Maximum Entropy Markov Model (MEMM), the classifier makes a single decision at a time, conditioned on evidence from observations and previous decisions
- A larger space of sequences is usually explored via search

Local Context

Decision Point

-3	-2	-1	0	+1
DT	NNP	VBD	???	???
The	Dow	fell	22.6	%

Features

W_0	22.6
W_{+1}	%
W_{-1}	fell
T_{-1}	VBD
T_{-1}-T_{-2}	NNP-VBD
hasDigit?	true
…	…

(Ratnaparkhi 1996; Toutanova et al. 2003, etc.)

Example: POS Tagging

- Scoring individual labeling decisions is no more complex than standard classification decisions
  - We have some assumed labels to use for prior positions
  - We use features of those and the observed data (which can include current, previous, and next words) to predict the current label

Local Context

Decision Point

-3	-2	-1	0	+1
DT	NNP	VBD	???	???
The	Dow	fell	22.6	%

Features

W_0	22.6
W_{+1}	%
W_{-1}	fell
T_{-1}	VBD
T_{-1}-T_{-2}	NNP-VBD
hasDigit?	true
…	…

(Ratnaparkhi 1996; Toutanova et al. 2003, etc.)

Example: POS Tagging

- POS tagging Features can include:
  - Current, previous, next words in isolation or together.
  - Previous one, two, three tags.
  - Word-internal features: word types, suffixes, dashes, etc.

Local Context

Decision Point

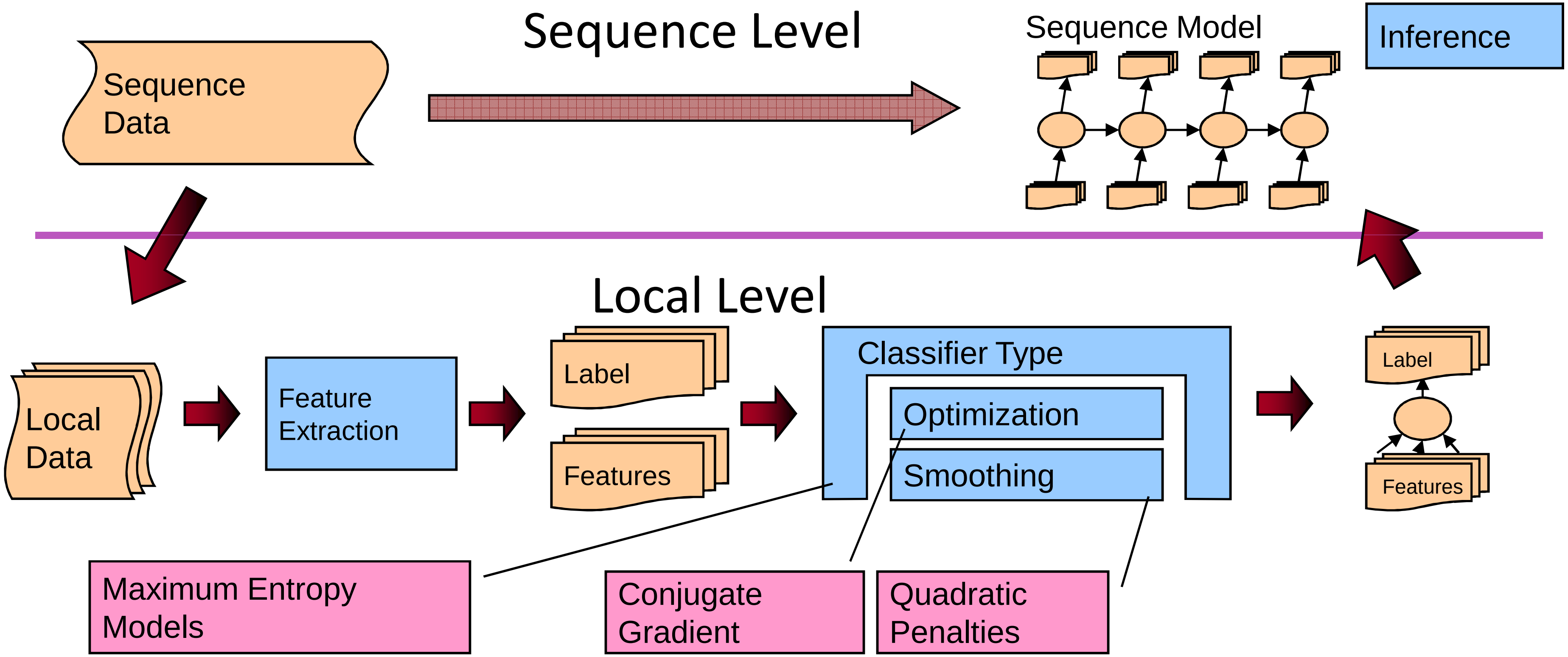
-3	-2	-1	0	+1
DT	NNP	VBD	???	???
The	Dow	fell	22.6	%

Features

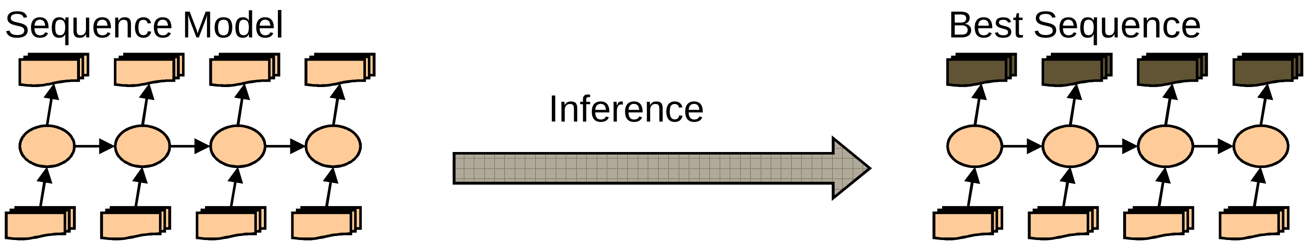
W_0	22.6
W_{+1}	%
W_{-1}	fell
T_{-1}	VBD
T_{-1}-T_{-2}	NNP-VBD
hasDigit?	true
…	…

(Ratnaparkhi 1996; Toutanova et al. 2003, etc.)

Inference in Systems

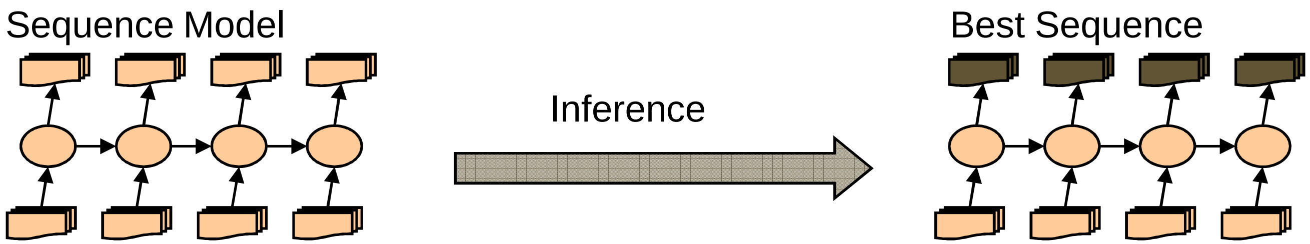


Greedy Inference



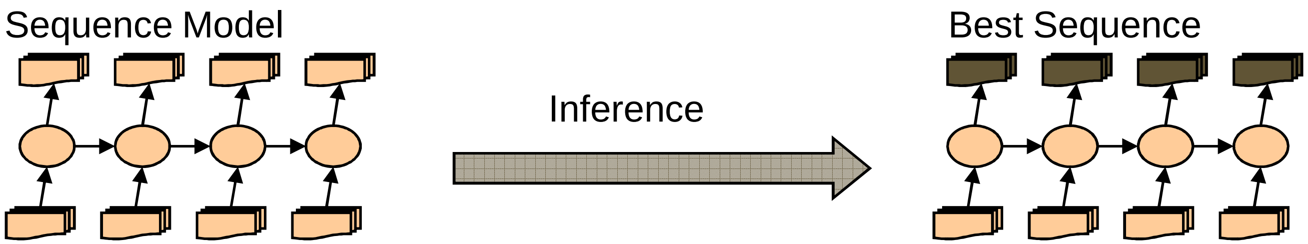
- Greedy inference:
  - We just start at the left, and use our classifier at each position to assign a label
  - The classifier can depend on previous labeling decisions as well as observed data
- Advantages:
  - Fast, no extra memory requirements
  - Very easy to implement
  - With rich features including observations to the right, it may perform quite well
- Disadvantage:
  - Greedy. We make commit errors we cannot recover from

Beam Inference



- Beam inference:
  - At each position keep the top k complete sequences.
  - Extend each sequence in each local way.
  - The extensions compete for the k slots at the next position.
- Advantages:
  - Fast; beam sizes of 3–5 are almost as good as exact inference in many cases.
  - Easy to implement (no dynamic programming required).
- Disadvantage:
  - Inexact: the globally best sequence can fall off the beam.

Viterbi Inference

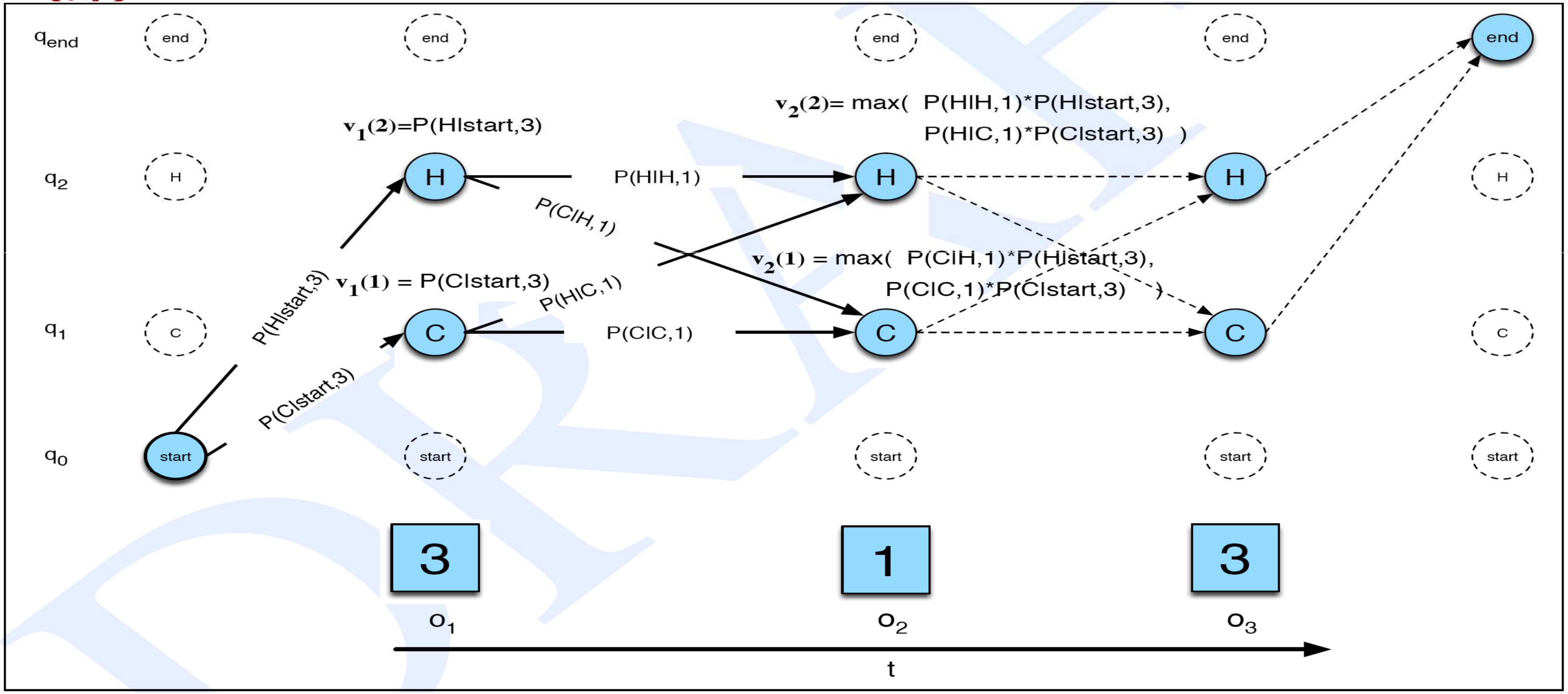


- Viterbi inference:
  - Dynamic programming or memoization.
  - Requires small window of state influence (e.g., past two states are relevant).
- Advantage:
  - Exact: the global best sequence is returned.
- Disadvantage:
  - Harder to implement long-distance state-state interactions (but beam inference tends not to allow long-distance resurrection of sequences anyway).

Viterbi Inference: J&M Ch. 6

- I'm punting on this … read J&M Ch. 5/6.
  - I'll do dynamic programming for parsing
- Basically, providing you only look at neighboring states, you can dynamic program a search for the optimal state sequence

Viterbi Inference: J&M Ch. 5/6



CRFs [Lafferty, Pereira, and McCallum 2001]

- Another sequence model: Conditional Random Fields (CRFs)
- A whole-sequence conditional model rather than a chaining of local models.

$$P(c \mid d, \lambda) = \frac{\exp \sum_i \lambda_i f_i(c, d)}{\sum_{c'} \exp \sum_i \lambda_i f_i(c', d)}$$

- The space of c' s is now the space of sequences
  - But if the features f_i remain local, the conditional sequence likelihood can be calculated exactly using dynamic programming
- Training is slower, but CRFs avoid causal-competition biases
- These (or a variant using a max margin criterion) are seen as the state-of-the-art these days … but in practice usually work much the same as MEMMs.

Maximum entropy sequence models

Maximum entropy Markov models (MEMMs) or Conditional Markov models

The Full Task of Information Extraction

The Full Task of Information Extraction

As a family of techniques:

Information Extraction = segmentation + classification + association + clustering

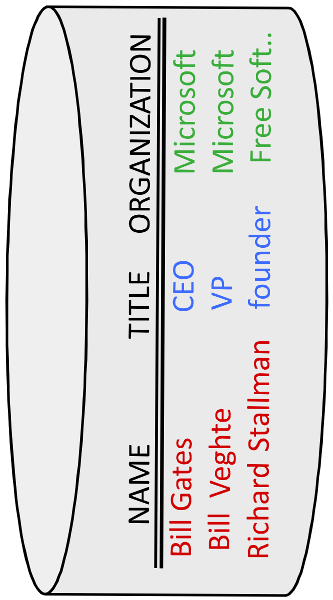
For years, Microsoft Corporation CEO Bill Gates railed against the economic philosophy of open-source software with Orwellian fervor, denouncing its communal licensing as a "cancer" that stifled technological innovation.

Now Gates himself says Microsoft will gladly disclose its crown jewels--the coveted code behind the Windows operating system--to select customers.

"We can be open source. We love the concept of shared source," said Bill Veghte, a Microsoft VP. "That's a super-important shift for us in terms of code access.“

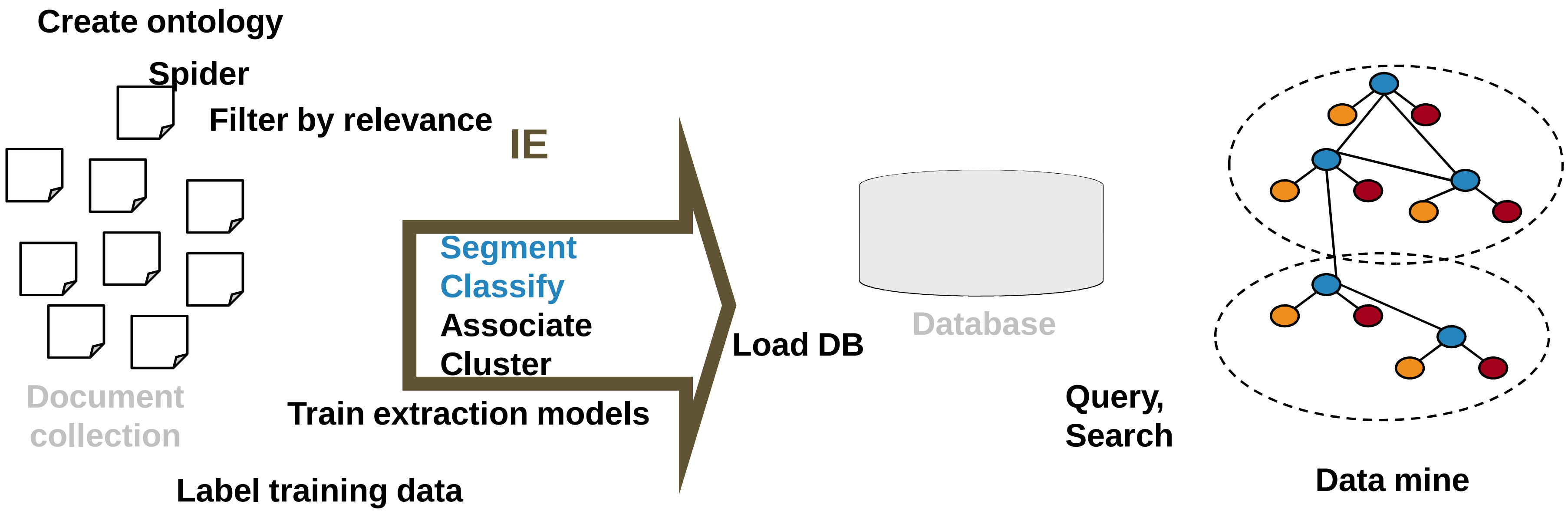
Richard Stallman, founder of the Free Software Foundation, countered saying…

Microsoft Corporation CEO Bill Gates
Gates Microsoft
Bill Veghte Microsoft VP
Richard Stallman founder Free Software Foundation



Slide by Andrew McCallum. Used with permission.

An Even Broader View



Slide by Andrew McCallum. Used with permission.

Landscape of IE Tasks:
Document Formatting

Text paragraphs without formatting

Astro Teller is the CEO and co-founder of BodyMedia. Astro holds a Ph.D. in Artificial Intelligence from Carnegie Mellon University, where he was inducted as a national Hertz fellow. His M.S. in symbolic and heuristic computation and B.S. in computer science are from Stanford University.

Grammatical sentences and some formatting & links

Dr. Steven Minton - Founder/CTO
Dr. Minton is a fellow of the American Association of Artificial Intelligence and was the founder of the Journal of Artificial Intelligence Research. Prior to founding Fetch, Minton was a faculty member at USC and a project leader at USC's Information Sciences Institute. A graduate of Yale University and Carnegie Mellon University, Minton has been a Principal Investigator at NASA Ames and taught at Stanford, UC Berkeley and USC.

- Press

Contact

- General information
- Directions maps

Non-grammatical snippets, rich formatting & links

Barto, Andrew G. (413) 545-2109 barto@cs.umass.edu CS276
Professor.
Computational neuroscience, reinforcement learning, adaptive motor control, artificial neural networks, adaptive and learning control, motor development.

Berger, Emery D. (413) 577-4211 emery@cs.umass.edu CS344
Assistant Professor.

Brock, Oliver (413) 577-0334 oli@cs.umass.edu CS246
Assistant Professor.

Clarke, Lori A. (413) 545-1328 clarke@cs.umass.edu CS304
Professor.
Software verification, testing, and analysis; software architecture and design.

Tables

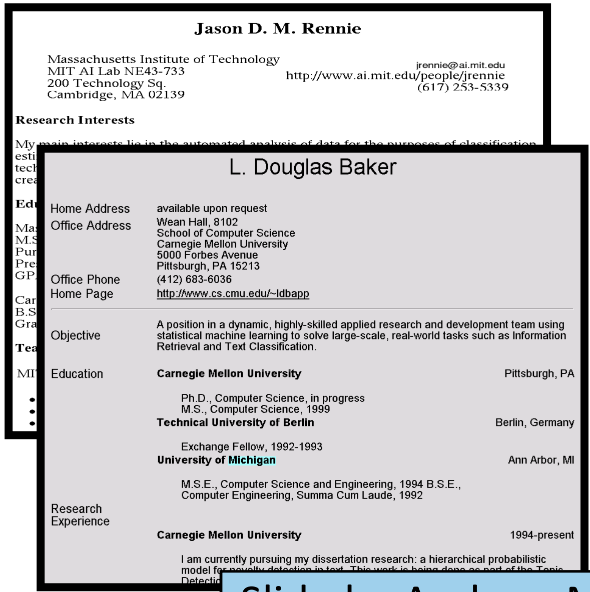
8:30 - 9:30 AM	**Invited Talk: Plausibility Measures: A General Approach for Representing Uncertainty** *Joseph Y. Halpern, Cornell University*				
9:30 - 10:00 AM	Coffee Break				
10:00 - 11:30 AM	Technical Paper Sessions:				
Cognitive Robotics	**Logic Programming**	**Natural Language Generation**	**Complexity Analysis**	**Neural Networks**	**Games**
739: A Logical Account of Causal and Topological Maps *Emilio Remolina and Benjamin Kuipers*	116: A-System: Problem Solving through Abduction *Marc Denecker, Antonis Kakas, and Bert Van Nuffelen*	758: Title Generation for Machine-Translated Documents *Rong Jin and Alexander G. Hauptmann*	417: Let's go Nats: Complexity of Nested Circumscription and Abnormality Theories *Marco Cadoli, Thomas Eiter, and Georg Gottlob*	179: Knowledge Extraction and Comparison from Local Function Networks *Kenneth McGarry, Stefan Wermter, and John MacIntyre*	71: Iterative Widening *Tristan Cazenave*
549: Online-Execution of ccGolog Plans *Henrik Grosskreutz*	131: A Comparative Study of Logic Programs with	246: Dealing with Dependencies between Content Planning and	470: A Perspective on Knowledge Compilation	258: Violation-Guided Learning for Constrained	353: Temporal Difference Learning Applied to a

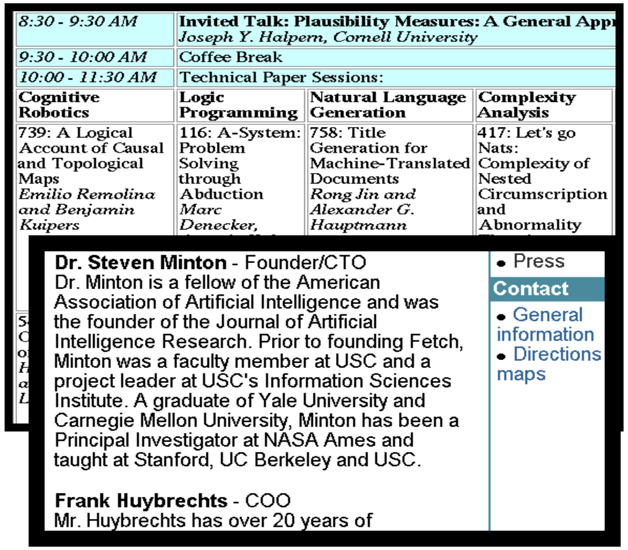
Slide by Andrew McCallum. Used with permission.

Christopher Manning

Landscape of IE Tasks
Intended Breadth of Coverage

Web site specific	Genre specific	Wide, non-specific
Formatting	Layout	Language
Amazon.com Book Pages	Resumes	University Names





Slide by Andrew McCallum. Used with permission.

Landscape of IE Tasks :
Complexity of entities/relations

Closed set

U.S. states

He was born in Alabama…

The big Wyoming sky…

Regular set

U.S. phone numbers

Phone: (413) 545-1323

The CALD main office is 412-268-1299

Complex pattern

U.S. postal addresses

University of Arkansas
P.O. Box 140
Hope, AR

Headquarters:
1128 Main Street, 4th Floor
Cincinnati, Ohio 45210

Ambiguous patterns, needing context and many sources of evidence

Person names

…was among the six houses sold by Hope Feldman that year.

Pawel Opalinski, Software Engineer at WhizBang Labs.

Slide by Andrew McCallum. Used with permission.

Landscape of IE Tasks:
Arity of relation

Jack Welch will retire as CEO of General Electric tomorrow. The top role at the Connecticut company will be filled by Jeffrey Immelt.

Single entity

Person: Jack Welch

Person: Jeffrey Immelt

Location: Connecticut

Binary relationship

Relation: Person-Title
Person: Jack Welch
Title: CEO

Relation: Company-Location
Company: General Electric
Location: Connecticut

N-ary record

Relation: Succession
Company: General Electric
Title: CEO
Out: Jack Welsh
In: Jeffrey Immelt

"Named entity" extraction

Slide by Andrew McCallum. Used with permission.

Association task = Relation Extraction

- Checking if groupings of entities are instances of a relation

1. Manually engineered rules
   - Rules defined over words/entites: “<company> located in <location>”
   - Rules defined over parsed text:
     - “((Obj <company>) (Verb located) (*) (Subj <location>))”

2. Machine Learning-based
   - Supervised: Learn relation classifier from examples
   - Partially-supervised: bootstrap rules/patterns from “seed” examples

Relation Extraction: Disease Outbreaks

May 19 1995, Atlanta -- The Centers for Disease Control and Prevention, which is in the front line of the world's response to the deadly Ebola epidemic in Zaire, is finding itself hard pressed to cope with the crisis…

Information Extraction System

Date	*Disease Name*	*Location*
Jan. 1995	Malaria	Ethiopia
July 1995	Mad Cow Disease	U.K.
Feb. 1995	Pneumonia	U.S.

Relation Extraction: Protein Interactions

“We show that CBF-A and CBF-C interact with each other to form a CBF-A-CBF-C complex and that CBF-B does not interact with CBF-A or CBF-C individually but that it associates with the CBF-A-CBF-C complex.“

$$\text{CBF-A} \xleftrightarrow[\text{complex}]{\text{interact}} \text{CBF-C}$$

$$\text{CBF-B} \xrightarrow[\text{associates}]{} \text{CBF-A-CBF-C complex}$$

Binary Relation Association as Binary Classification

Christos Faloutsos conferred with Ted Senator, the KDD 2003 General Chair.

Person Person Role

Person-Role (Christos Faloutsos, KDD 2003 General Chair) → NO

Person-Role (Ted Senator, KDD 2003 General Chair) → YES

Christopher Manning

Resolving coreference (both within and across documents)

John Fitzgerald Kennedy was born at 83 Beals Street in Brookline, Massachusetts on Tue 29, 1917, at 3:00 pm,[7] the second son of Joseph P. Kennedy, Sr., and Rose Fitzgerald; R turn, was the eldest child of John "Honey Fitz" Fitzgerald, a prominent Boston political fi was the city's mayor and a three-term member of Congress. Kennedy lived in Brookline years and attended Edward Devotion School, Noble and Greenough Lower School, and the Dexter School, through 4th grade. In 1927, the family moved to 5040 Independence Avenue in Bronx, New York City; two years later, they moved to 294 Pondfield Road in Bronxville, N where Kennedy was a member of Scout Troop 2 (and was the first Boy Scout to become President).[8] Kennedy spent summers with his family at their home in Hyannisport, Massachusetts, and Christmas and Easter holidays with his family at their winter home i Beach, Florida. For the 5th through 7th grade, Kennedy attended Riverdale Country School, a private school for boys. For 8th grade in September 1930, the 13-year old Kennedy attended Canterbury School in New Milford, Connecticut.

Rough Accuracy of Information Extraction

Information type	Accuracy
Entities	90-98%
Attributes	80%
Relations	60-70%
Events	50-60%

- Errors cascade (error in entity tag → error in relation extraction)
- These are very rough, actually optimistic, numbers
  - Hold for well-established tasks, but lower for many specific/novel IE tasks

The Full Task of Information Extraction