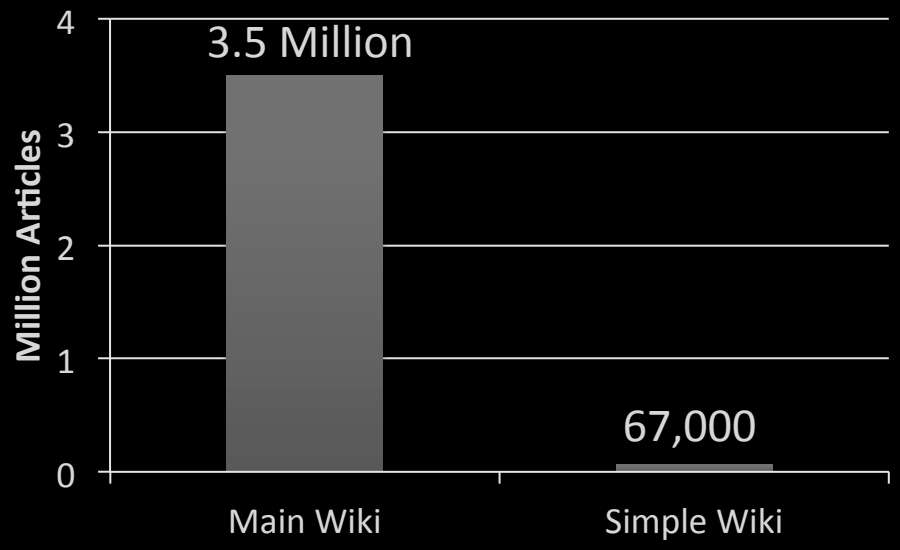
EE 570 Paper Critiques and Implementation

Ryan Miller
November 29, 2012

WikiSimple: Automatic Simplification of Wikipedia Articles

- Motivation



Woodsend, K. and Lapata, M. 2011. WikiSimple: Automatic Simplification of Wikipedia Articles. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, 927- 932.

WikiSimple: Automatic Simplification of Wikipedia Articles

- Goal
  - Automate the creation of Simple Wiki articles

Main Wiki

Owls are a group of birds that belong to the order *Strigiformes*, constituting 200 bird of prey species. Owls hunt mostly small mammals, insects, and other birds, though some species specialize in hunting fish.

Simple Wiki

An owl is a bird. There are about 200 kinds of owls. Owls' prey may be birds, large insects (such as crickets), small reptiles (such as lizards) or small mammals (such as mice, rats, and rabbits).

WikiSimple: Automatic Simplification of Wikipedia Articles

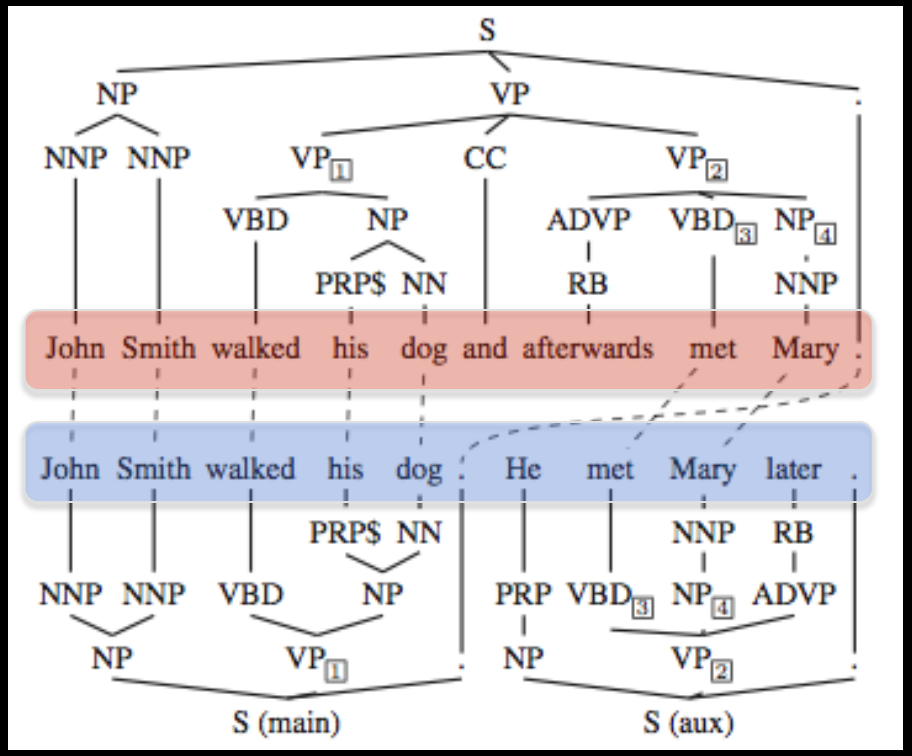
- Techniques
  - Quasi-synchronous grammar restructuring
    - Set of rules to follow to make new phrases
    - Learned from the revision history of Simple Wiki articles
  - Integer linear programming (ILP)
    - Choosing the best phrases to use from the Main Wiki article

WikiSimple: Automatic Simplification of Wikipedia Articles

- Quasi-synchronous grammar restructuring

Source sentence

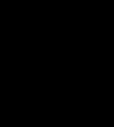
Auxiliary sentences



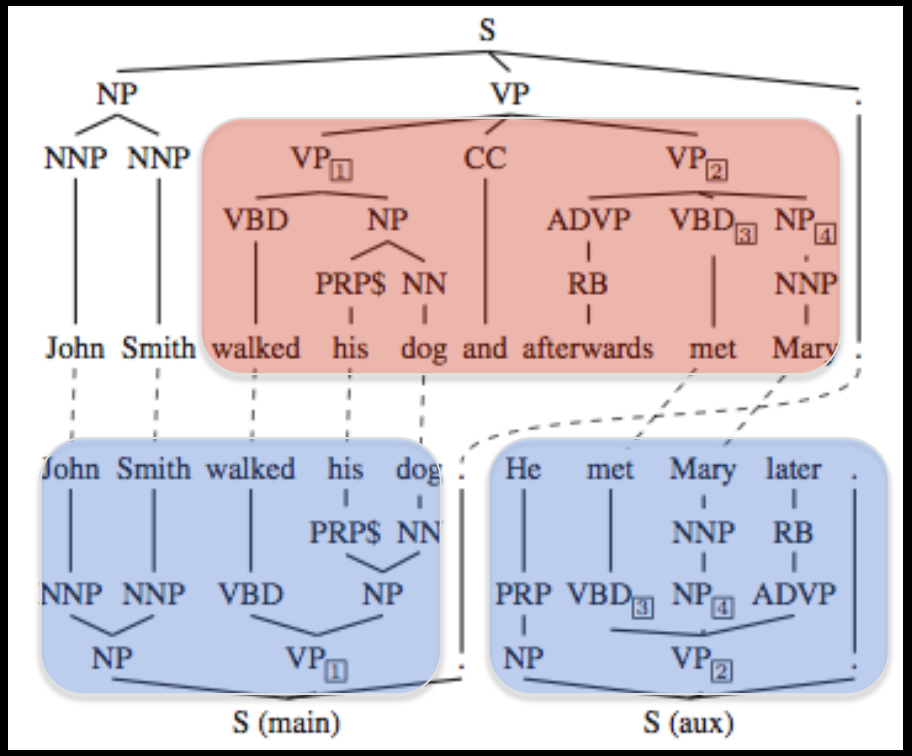
WikiSimple: Automatic Simplification of Wikipedia Articles

- Quasi-synchronous grammar restructuring

< VP_1, VP_2, S>



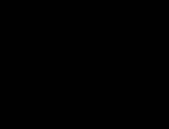
< [NP VP_1]. [NP VP_2].>



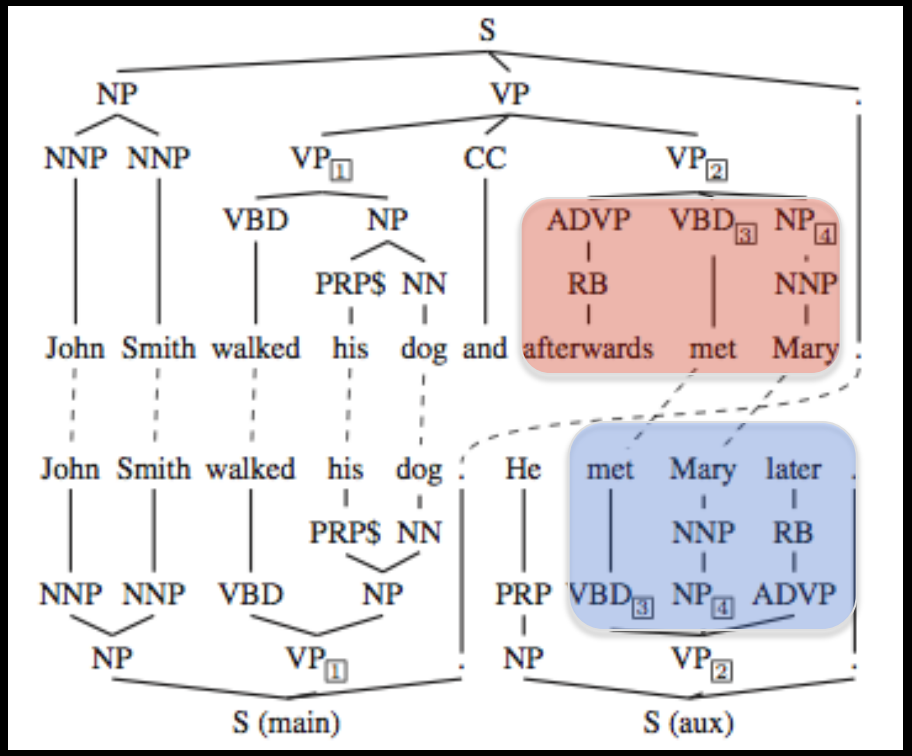
WikiSimple: Automatic Simplification of Wikipedia Articles

- Quasi-synchronous grammar restructuring

< [[ADVP *afterwards]* VBD NP]>



< [VBD NP [ADVP *later*]]>



WikiSimple: Automatic Simplification of Wikipedia Articles

- Integer Linear Programming (ILP)
  - Need to solve an equation when some or all of the variables are unknown
  - NP-hard

WikiSimple: Automatic Simplification of Wikipedia Articles

- Binary ILP: $x_i \in \{0,1\}$

$$\max_{x} \quad \sum_{i \in \mathcal{P}} (f_i + g_i) x_i + h_w + h_{sy} \qquad \text{(1a)}$$

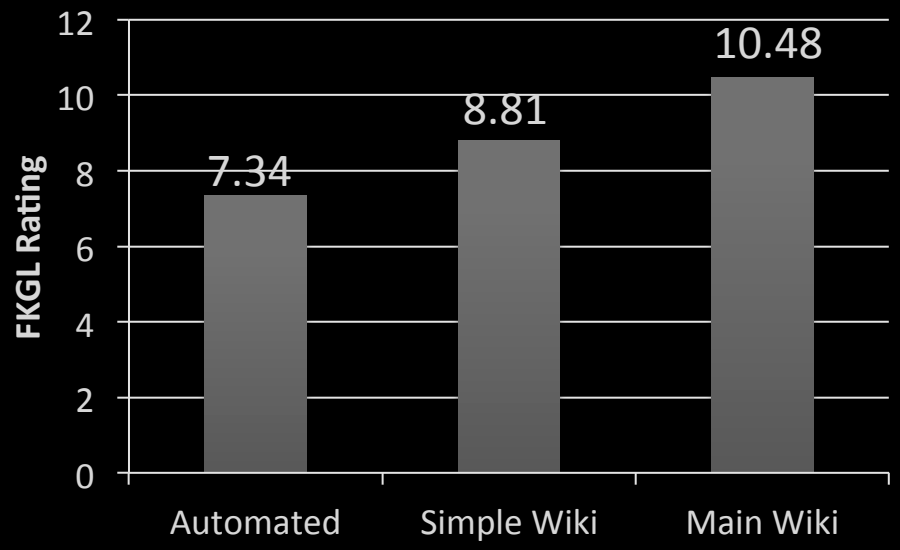
Salience Score

Rewrite Penalty

Approximation of FKGL index

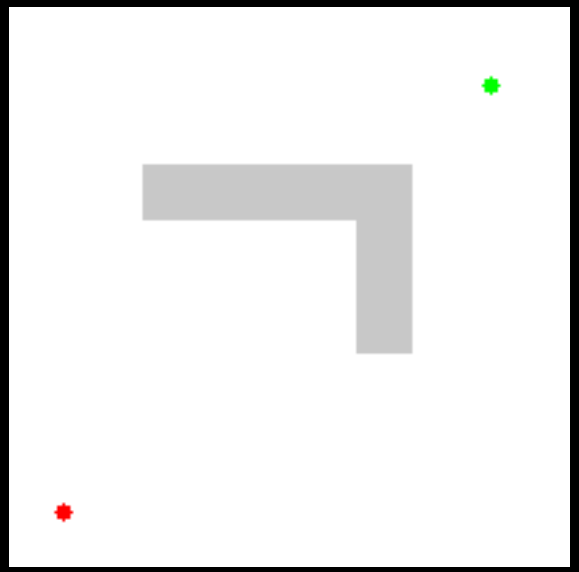
WikiSimple: Automatic Simplification of Wikipedia Articles

- Results
  - Generated 1,654 Simple Wiki articles
  - Average FKGL rating



Online Graph Pruning for Pathfinding on Grid Maps

- Motivation

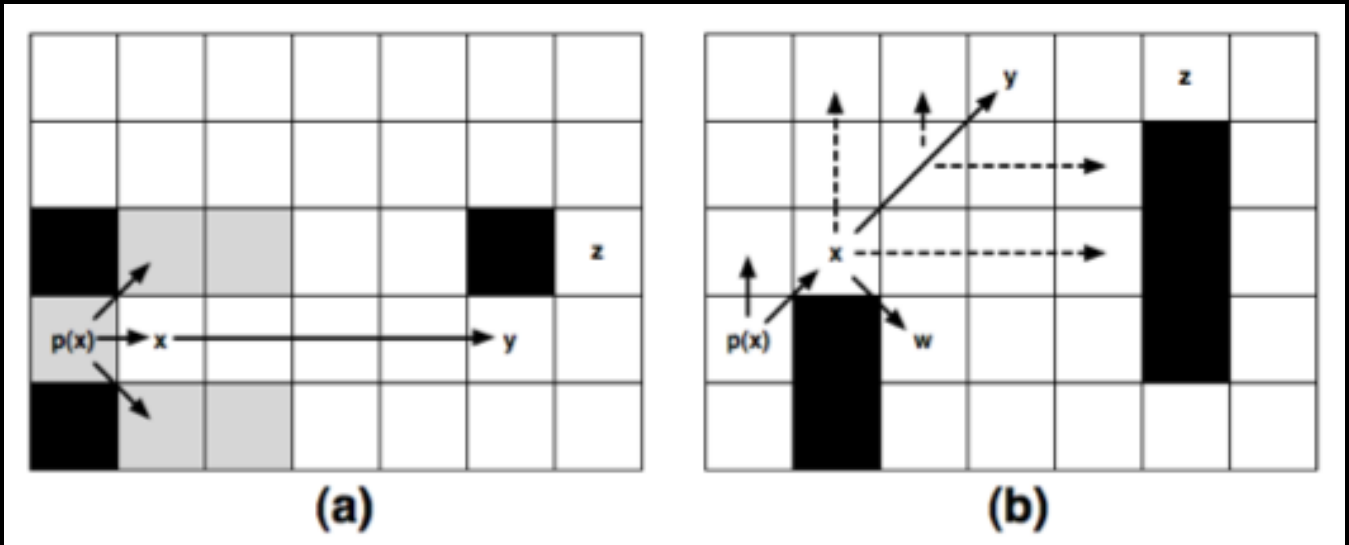


"A* search algorithm," Wikipedia

Harabor, D. and Grastien, A. 2011. Online Graph Pruning for Pathfinding on Grid Maps. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, 1114-1119.

Online Graph Pruning for Pathfinding on Grid Maps

- Technique
  - Jump Points



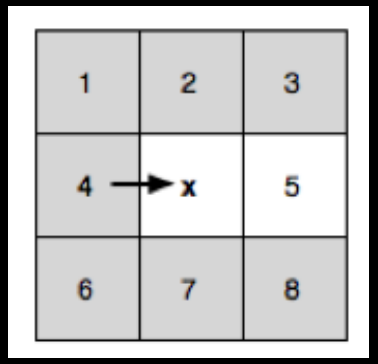
Online Graph Pruning for Pathfinding on Grid Maps

- Technique
  - Jump Points
  - Eliminate node N from neighbors(X) that satisfies the dominance constraint:

$$\text{cost}(\langle p(X), \ldots, N \rangle \setminus X) \leq \text{cost}(\langle p(X), X, N \rangle)$$

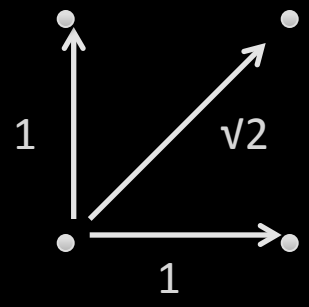
Predecessor(X)

Set minus X



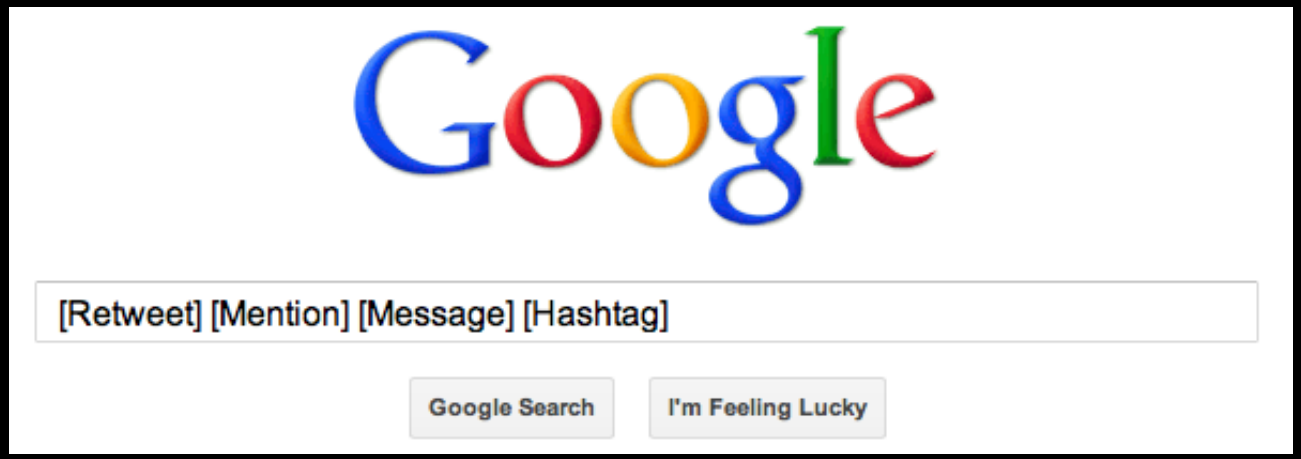
Online Graph Pruning for Pathfinding on Grid Maps

- Results
  - Faster than A*
    - "by an order of magnitude"
  - As fast as Heuristic Pathfinding Algorithms (HPA)
    - But always optimal
- Caveat
  - Only works for uniform cost grid maps



Improving Twitter Retrieval by Exploiting Structural Information

- Motivation



RT @ladygaga OMG this is my fav song #yolo

Luo, Z., Osborne, M., Petrovic, S., & Wang, T. 2012. Improving Twitter Retrieval by Exploiting Structural Information. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, 648-654.

Improving Twitter Retrieval by Exploiting Structural Information

- Goal
  - Parse a tweet into **tokens**
  - Categorize the tokens into structural blocks – **Twitter Building Blocks** (TBBs)

Input						
RT @CBCNews Tony Curtis dies at 85 http://bit.ly/dlS3						
Output						
RT @CBCNews	Tony	Curtis	dies	at	85	http://bit.ly/dlS3
	PN	**PN**	**V**	**P**	**#**	
Re-tweet	**Message**					**URL**

Improving Twitter Retrieval by Exploiting Structural Information

- Techniques
  - Parse a tweet into tokens
    - Tweet NLP and Part-of-speech Tagger
  - Manually tag ~2,000 tweets for training, development, and test sets

Improving Twitter Retrieval by Exploiting Structural Information

- Types of TBB structures
  - Hashtag (‘#’<phrase>)
  - Mention (‘@’<user>)
  - Re-tweet (‘RT’ <Mention>)
  - URL (‘http://’<url>)
  - Commentary (opinion or feelings)
  - Message (sentence content)
  - Other

Improving Twitter Retrieval by Exploiting Structural Information

- Techniques
  - *Conditional Random Field* features considered
    - Token type
    - Part of speech
    - Token length
    - Twitter orthography (Hashtags begin with '#', etc.)

TBB Structure	Per.(%)	TBB Structure	Per.(%)
MSG	30.25	TAG MSG	1.55
MET MSG	20.70	TAG MSG URL	1.20
MSG URL	18.40	RWT MSG URL	0.95
OTHERS	13.20	COM RWT MSG	0.85
COM URL	4.10	MET MSG URL	0.85
MSG TAG	2.65	MSG MET MSG	0.70
MSG URL TAG	2.10	RWT MSG TAG	0.70
RWT MSG	1.75		

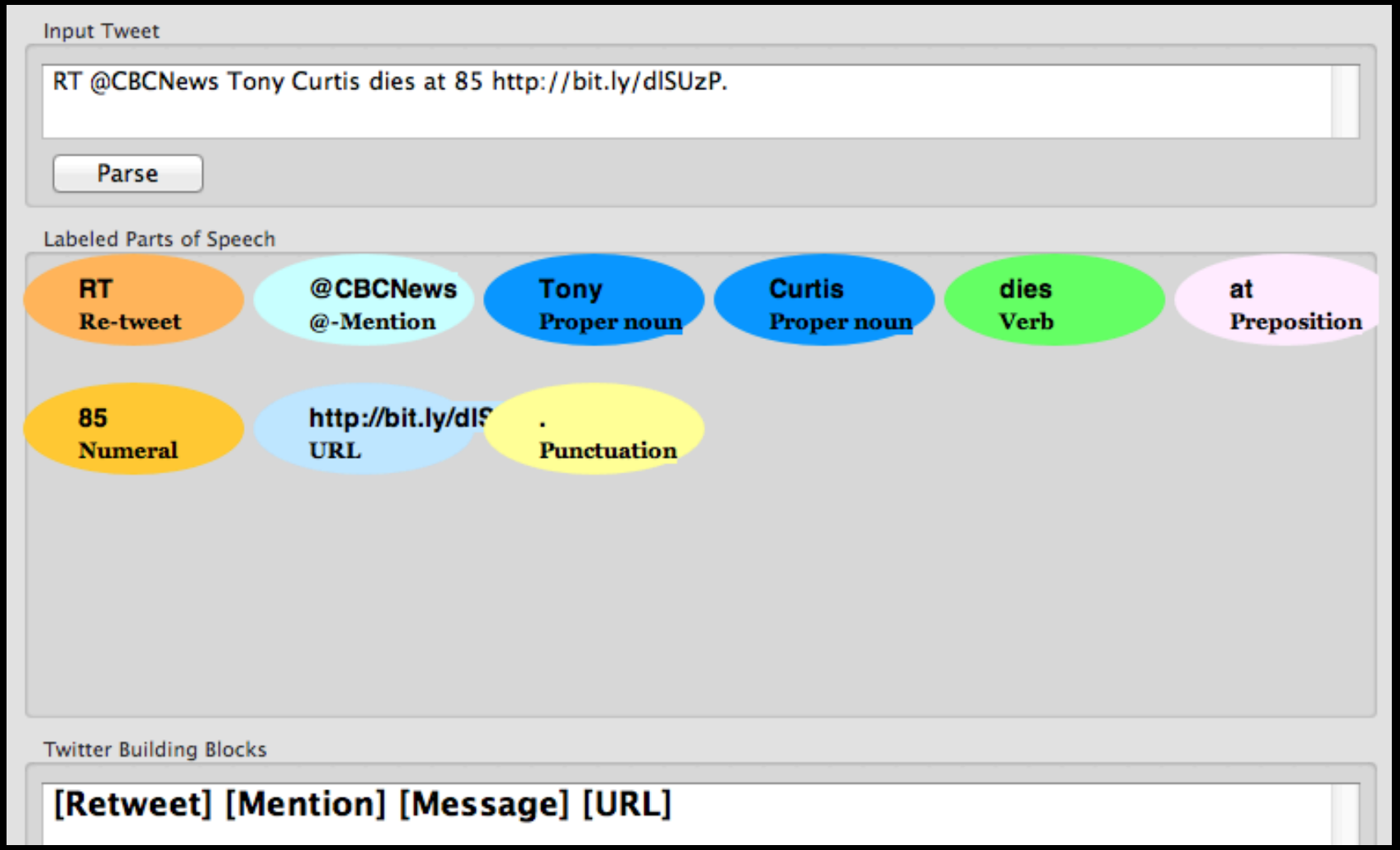
Improving Twitter Retrieval by Exploiting Structural Information

- My implementation
  - TweetNLP for tokenizing tweets
  - Heuristic for categorizing tweets
    - Look at previous token
    - Maintain list of structures

- Demo 1: Parsing a single tweet

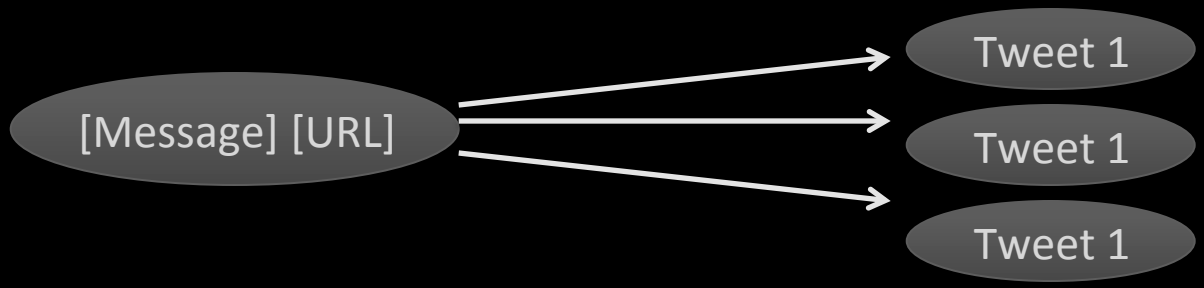
RT @CBCNews Tony Curtis dies at 85 http://bit.ly/dlSUzP.

Improving Twitter Retrieval by Exploiting Structural Information



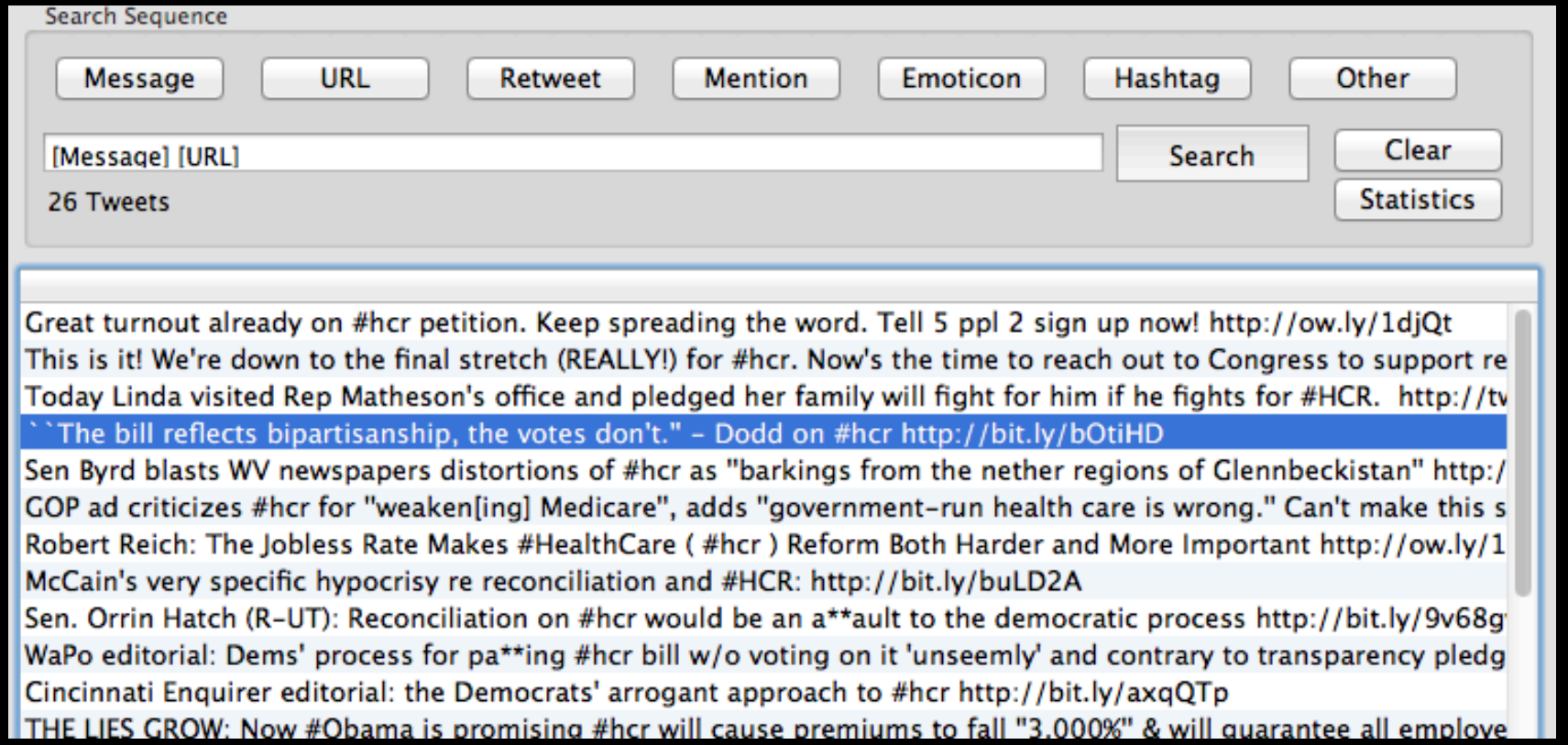
Improving Twitter Retrieval by Exploiting Structural Information

- My implementation
  - Data set of 1,340 tweets
  - Treat string of TBBs as "key" to hash table



- Demo 2: Retrieving tweets

Improving Twitter Retrieval by Exploiting Structural Information



Improving Twitter Retrieval by Exploiting Structural Information

- Results
  - Most popular structures

Table 1: Tweet Statistics (1,340 tweets)

TBB Structure	Percentage	Number
Message	99.93 %	1,339 tweets
Hashtag	87.84	1,177
URL	27.61	370
Mention	25.6	343
Re-tweet	16.94	227
Other	7.09	95
Emoticon	1.42	19

Improving Twitter Retrieval by Exploiting Structural Information

- Results
  - Most popular patterns of structures

Table 2: Tweet Statistics (1,340 tweets)

TBB Pattern	Percentage	Number
[Message] [Hashtag]	41.34 %	554 tweets
[Message] [URL] [Hashtag]	9.03	121
[Hashtag] [Message]	4.85	65
[Message]	3.96	53
[Mention] [Message] [Hashtag]	3.06	41
[Retweet] [Mention] [Retweet] [Message] [Hashtag]	2.31	31
[Message] [URL]	1.64	26

Questions?

EE 570 Paper Critiques and Implementation

Ryan Miller
November 29, 2012