Computational Linguistics

10A

CSC 2501 / 485
Fall 2018

Assignment Project Exam Help

10A. Log-Likelihood Dependency Parsing

https://powcoder.com

Add WeChat powcoder

Gerald Penn
Department of Computer Science, University of Toronto

Based on slides by Yuji Matsumoto, Dragomir Radev, David Smith and Jason Eisner

Copyright © 2017 Gerald Penn. All rights reserved.

Word Dependency Parsing

Raw sentence

He reckons the current account deficit will narrow to only 1.8 billion in September.

Part-of-speech tagging

POS-tagged sentence

He reckons the current account deficit will narrow to only 1.8 billion in September.

PRP VBZ DT JJ NN NN MD VB TO RB CD CD IN NNP .

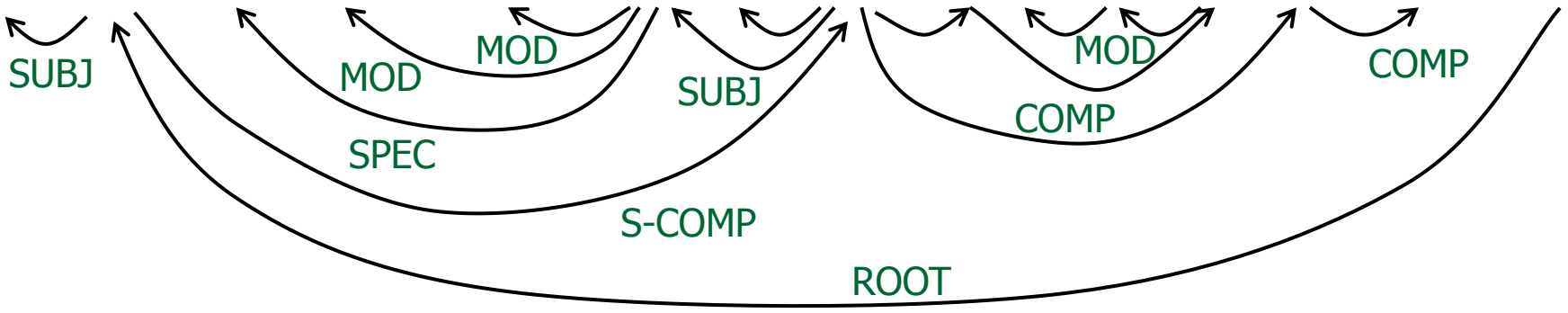
Assignment Project Exam Help
https://powcoder.com

Word dependency parsing

Add WeChat powcoder

Word dependency parsed sentence

He reckons the current account deficit will narrow to only 1.8 billion in September .



slide adapted from Yuji Matsumoto

Shift-Reduce Type Algorithms

Assignment Project Exam Help

- Data structures:
  - Stack $[\ldots, w_i]_S$ of partially processed tokens
  - Queue $[w_j, \ldots]_Q$ of remaining input tokens
- Parsing actions built from atomic actions:
  - Adding arcs ($w_i \rightarrow w_j$, $w_i \leftarrow w_j$)
  - Stack and queue operations
- Left-to-right parsing in $O(n)$ time
- Restricted to projective dependency graphs

https://powcoder.com

Add WeChat powcoder

Yamada's Algorithm

- Three parsing actions:

$$\text{Shift}\quad \frac{[\ldots]_S \quad [w_i, \ldots]_Q}{[\ldots, w_i]_S \quad [\ldots]_Q}$$

$$\text{Left}\quad \frac{[\ldots, w_i, w_j]_S \quad [\ldots]_Q}{[\ldots, w_i]_S \quad [\ldots]_Q \quad w_i \rightarrow w_j}$$

$$\text{Right}\quad \frac{[\ldots, w_i, w_j]_S \quad [\ldots]_Q}{[\ldots, w_j]_S \quad [\ldots]_Q \quad w_i \leftarrow w_j}$$

- Algorithm variants:
  - Originally developed for Japanese (strictly head-final) with only the Shift and Right actions [Kudo and Matsumoto 2002]
  - Adapted for English (with mixed headedness) by adding the Left action [Yamada and Matsumoto 2003]
  - Multiple passes over the input give time complexity $O(n^2)$

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Nivre's Algorithm

- Four parsing actions:

Assignment Project Exam Help

$$\text{Shift} \quad \frac{[\ldots]_S \quad [w_i, \ldots]_Q}{[\ldots, w_i]_S \quad [\ldots]_Q}$$

$$\text{Reduce} \quad \frac{[\ldots, w_i]_S \quad [\ldots]_Q \quad \exists w_k : w_k \rightarrow w_i}{[\ldots]_S \quad [\ldots]_Q}$$

https://powcoder.com

$$\text{Left-Arc}_r \quad \frac{[\ldots, w_i]_S \quad [w_j, \ldots]_Q \quad \neg\exists w_k : w_k \rightarrow w_i}{[\ldots]_S \quad [w_j, \ldots]_Q \quad w_i \overset{r}{\leftarrow} w_j}$$

$$\text{Right-Arc}_r \quad \frac{[\ldots, w_i]_S \quad [w_j, \ldots]_Q \quad \neg\exists w_k : w_k \rightarrow w_j}{[\ldots, w_i, w_j]_S \quad [\ldots]_Q \quad w_i \overset{r}{\rightarrow} w_j}$$

Add WeChat powcoder

- Characteristics:
  - Integrated labeled dependency parsing
  - Arc-eager processing of right-dependents
  - Single pass over the input gives time complexity $O(n)$

Example

Assignment Project Exam Help

https://powcoder.com

[root]$_S$ [Economic news had little effect on financial markets .]$_Q$

Add WeChat powcoder

Example

Assignment Project Exam Help

https://powcoder.com

[root Economic]$_S$ [news had little effect on financial markets .]$_Q$

Add WeChat powcoder

Shift

Example

Assignment Project Exam Help

https://powcoder.com

[root]$_S$ Economic [news had little effect on financial markets .]$_Q$

Add WeChat powcoder

Left-Arc$_{nmod}$

Example

Assignment Project Exam Help

https://powcoder.com

[root Economic news]$_S$ [had little effect on financial markets .]$_Q$

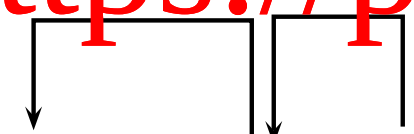
Add WeChat powcoder

Shift

Example

Assignment Project Exam Help

https://powcoder.com



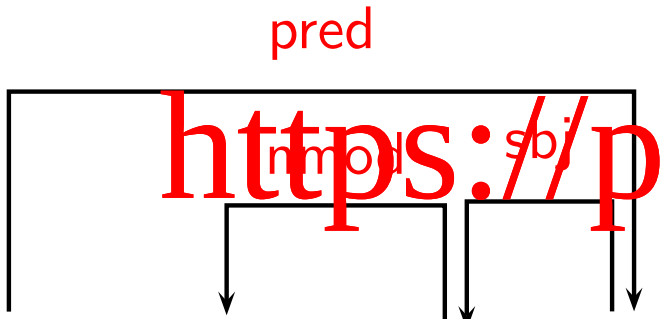
[root]$_S$ Economic news [had little effect on financial markets .]$_Q$

Add WeChat powcoder

Left-Arc$_{sbj}$

Example

Assignment Project Exam Help



https://powcoder.com

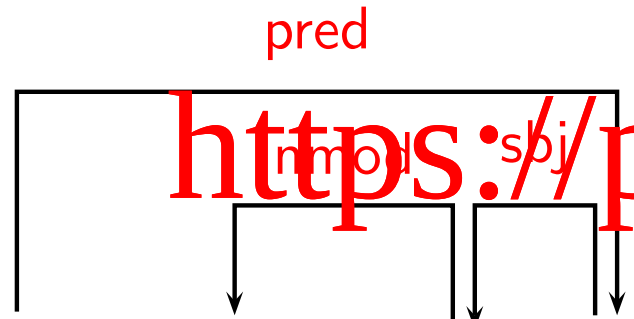
[root Economic news had]$_S$ [little effect on financial markets .]$_Q$

Add WeChat powcoder

Right-Arc$_{pred}$

Example

Assignment Project Exam Help



https://powcoder.com

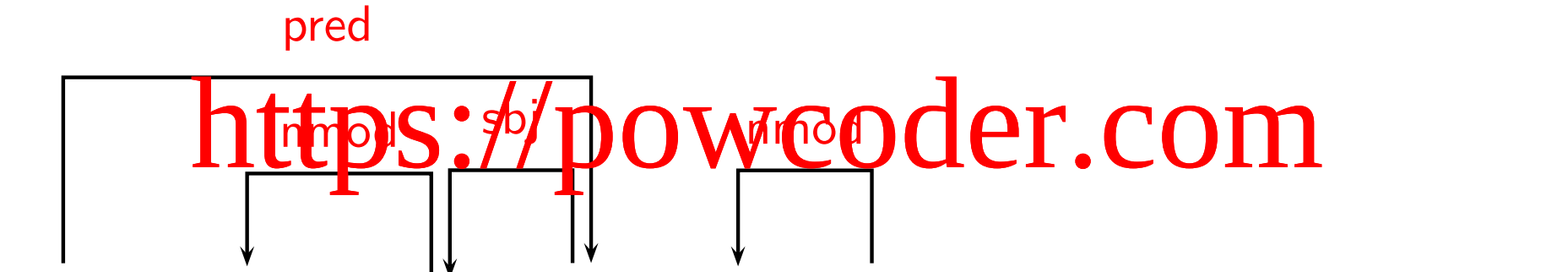
[root Economic news had little]$_S$ [effect on financial markets .]$_Q$

Add WeChat powcoder

Shift

Example

Assignment Project Exam Help



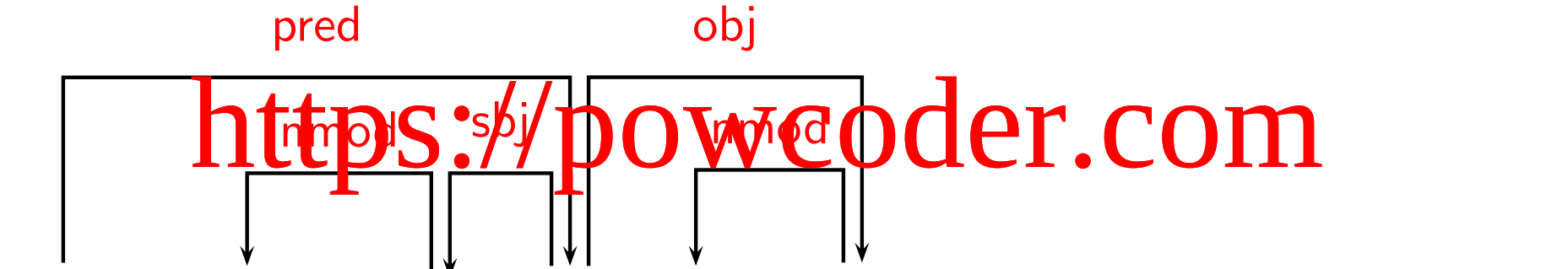
https://powcoder.com

Add WeChat powcoder

Left-Arc$_{nmod}$

Example

Assignment Project Exam Help



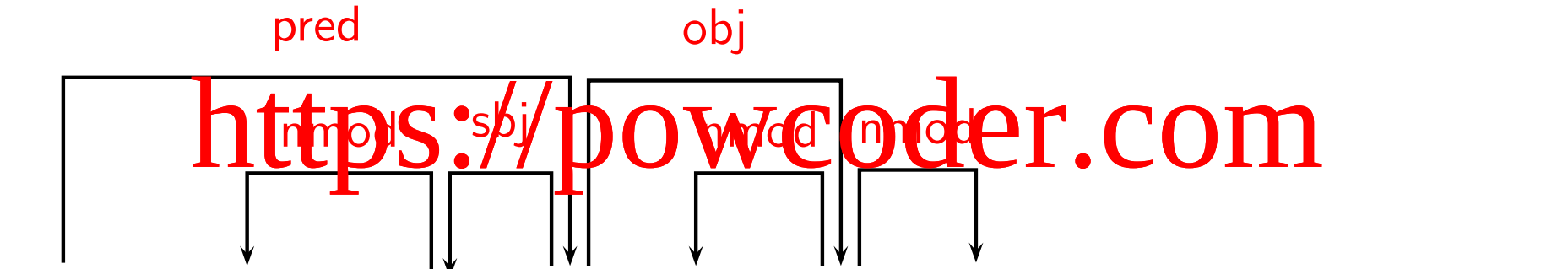
https://powcoder.com

Add WeChat powcoder

Right-Arc$_{obj}$

Example

Assignment Project Exam Help



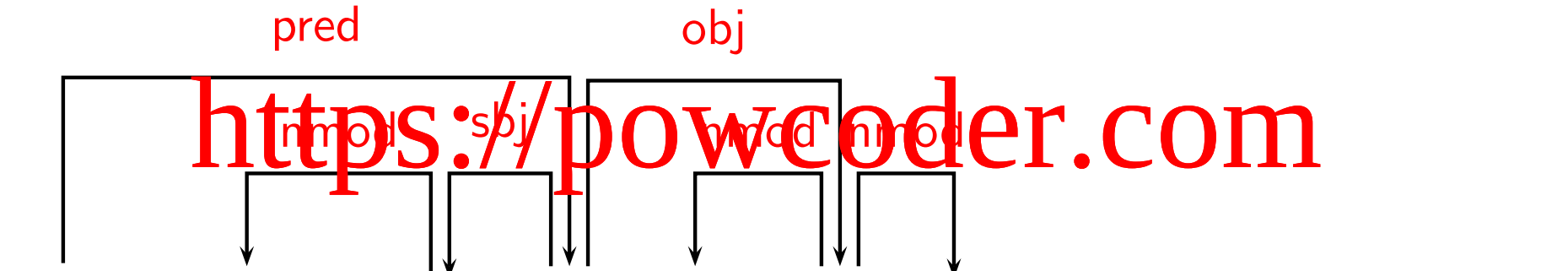
https://powcoder.com

Add WeChat powcoder

Right-Arc$_{nmod}$

Example

Assignment Project Exam Help



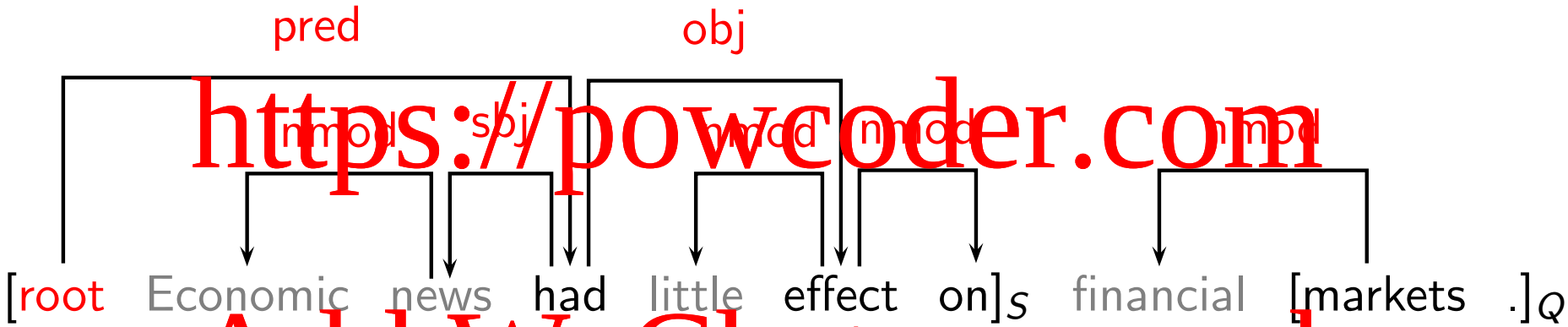
https://powcoder.com

Add WeChat powcoder

Shift

Example

Assignment Project Exam Help



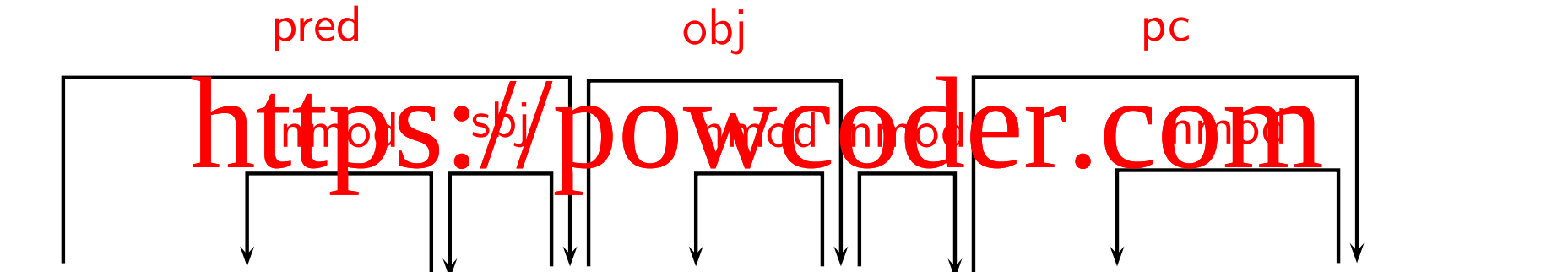
https://powcoder.com

Add WeChat powcoder

Left-Arc$_{nmod}$

Example

Assignment Project Exam Help



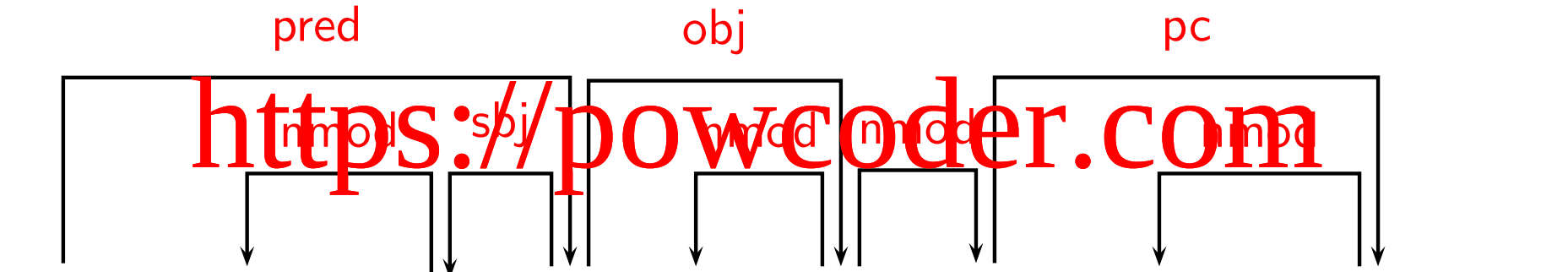
https://powcoder.com

Add WeChat powcoder

Right-Arc$_{pc}$

Example

Assignment Project Exam Help



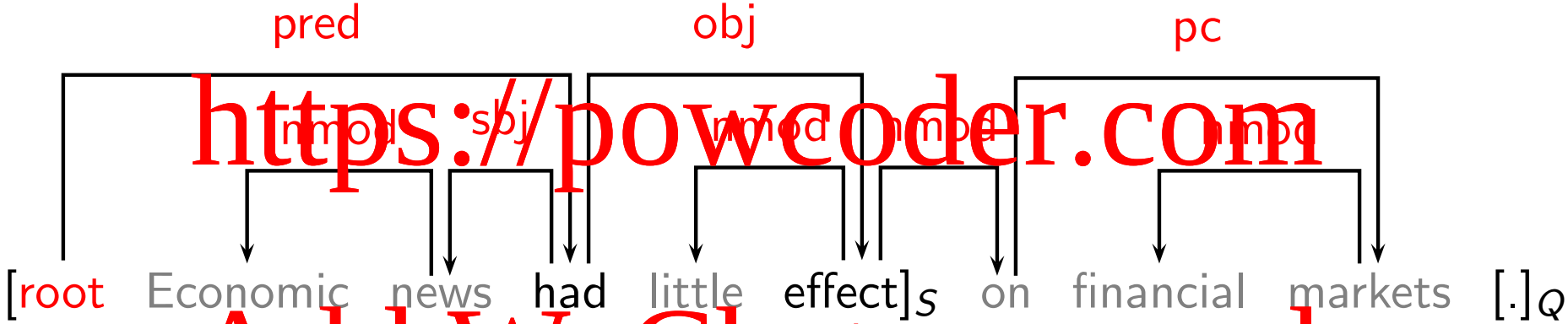
https://powcoder.com

Add WeChat powcoder

Reduce

Example

Assignment Project Exam Help



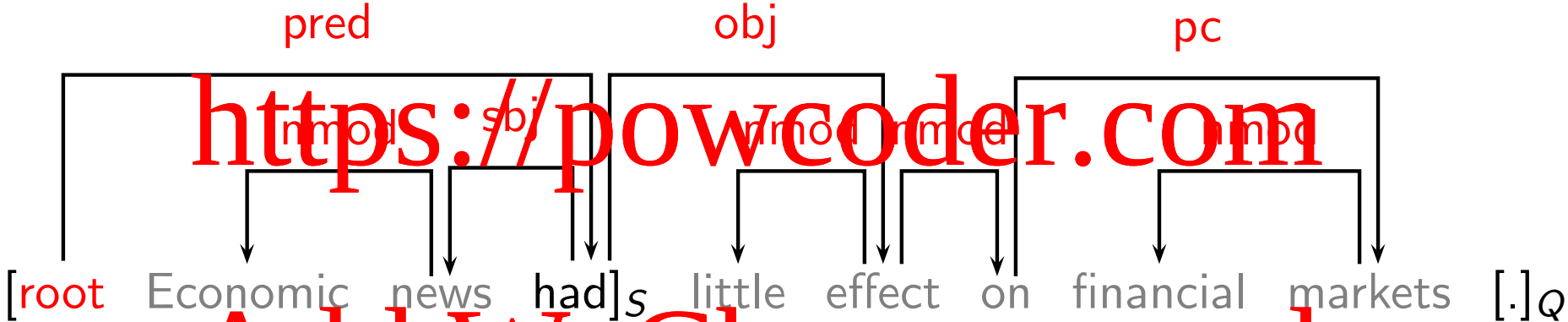
https://powcoder.com

Add WeChat powcoder

Reduce

Example

Assignment Project Exam Help



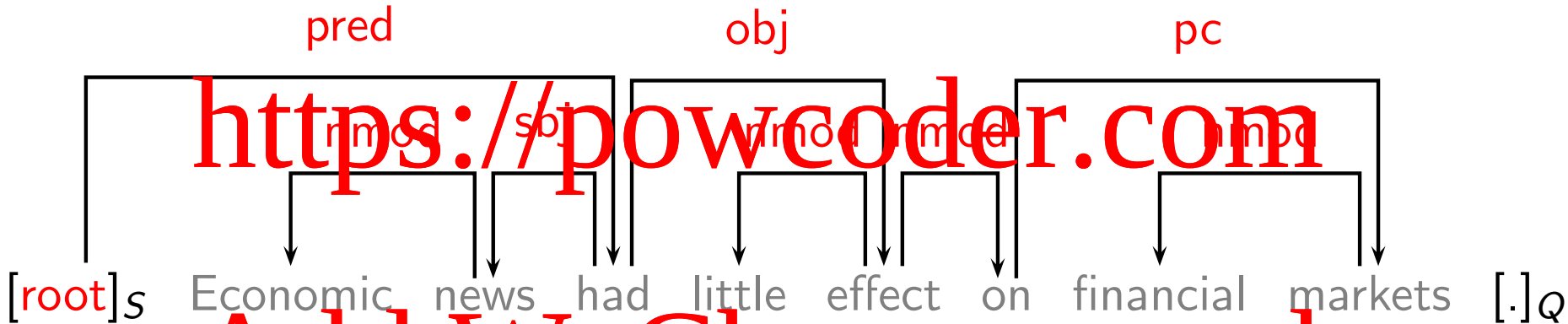
https://powcoder.com

Add WeChat powcoder

Reduce

Example

Assignment Project Exam Help



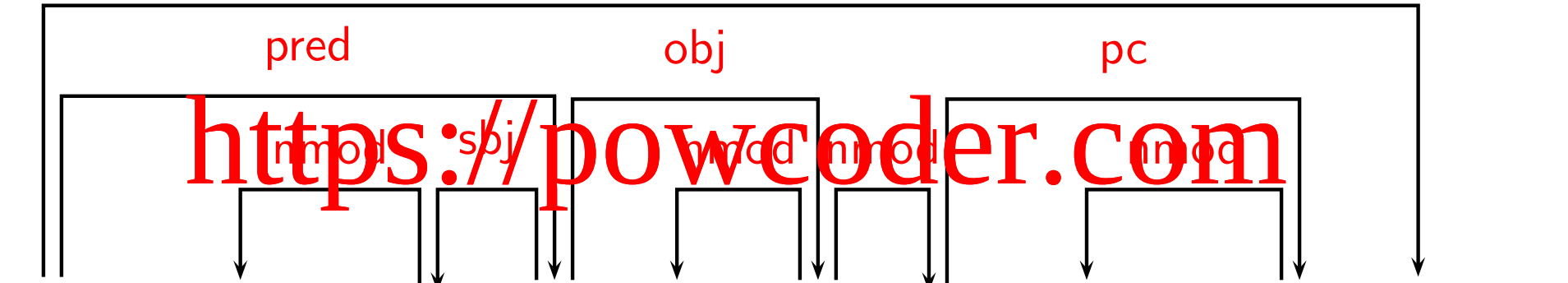
https://powcoder.com

Add WeChat powcoder

Reduce

Example

Assignment Project Exam Help
https://powcoder.com



Add WeChat powcoder

Right-Arc$_p$

Classifier-Based Parsing

Assignment Project Exam Help

- Data-driven deterministic parsing:
  - Deterministic parsing requires an oracle.
  - An oracle can be approximated by a classifier.
  - A classifier can be trained using treebank data.
- Learning methods:
  - Support vector machines (SVM)
    [Kudo and Matsumoto 2002, Yamada and Matsumoto 2003, Isozaki et al. 2004, Cheng et al. 2004, Nivre et al. 2006]
  - Memory-based learning (MBL)
    [Nivre et al. 2004, Nivre and Scholz 2004]
  - Maximum entropy modeling (MaxEnt)
    [Cheng et al. 2005]

https://powcoder.com

Add WeChat powcoder

Feature Models

- Learning problem:
  - Approximate a function from parser states, represented by feature vectors to parser actions, given a training set of gold standard derivations.
- Typical features:
  - Tokens:
    - Target tokens
    - Linear context (neighbors in S and Q)
    - Structural context (parents, children, siblings in G)
  - Attributes:
    - Word form (and lemma)
    - Part-of-speech (and morpho-syntactic features)
    - Dependency type (if labeled)
    - Distance (between target tokens)

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Great ideas in NLP: Log-linear models

(Berger, della Pietra, della Pietra 1996; Darroch & Ratcliff 1972)

- In the beginning, we used generative models.

$$p(A) * p(B \mid A) * p(C \mid A,B) * p(D \mid A,B,C) * \ldots$$

each choice depends on a *limited* part of the history

but which dependencies to allow?
what if they're all worthwhile?

$p(D \mid A,B,C)$?
$p(D \mid A,B,C)$?
$\ldots p(D \mid A,B) * p(C \mid A,B,D)$?

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Great ideas in NLP: Log-linear models

(Berger, della Pietra, della Pietra 1996; Darroch & Ratcliff 1972)

- In the beginning, we used generative models.

  p(A) * p(B | A) * p(C | A,B) * p(D | A,B,C) * …

  which dependencies to allow? (given limited training data)

- Solution: Log-linear (max-entropy) modeling

  (1/Z) * Φ(A) * Φ(B,A) * Φ(C,A) * Φ(C,B) * Φ(D,A,B) * Φ(D,B,C) * Φ(D,A,C) * …

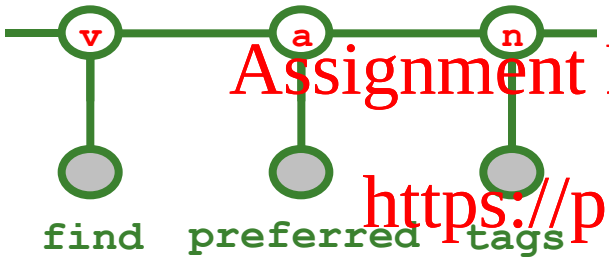
  …throw them all in!

  - Features may interact in arbitrary ways
  - **Iterative scaling** keeps adjusting the feature weights until the model agrees with the training data.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

How about structured outputs?

- Log-linear models great for n-way classification
- Also good for predicting sequences



but to allow fast dynamic programming, only use **n-gram** features

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

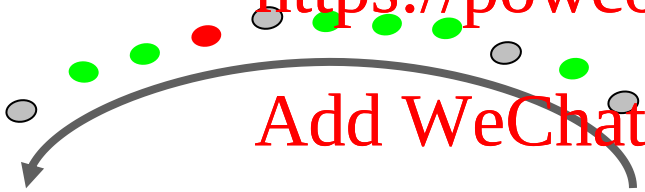
- Also good for dependency parsing

but to allow fast dynamic programming or MST parsing, only use **single-edge** features

Edge-Factored Parsers (McDonald et al. 2005)

- Is this a good edge?

Assignment Project Exam Help
yes, lots of green ...
https://powcoder.com
Add WeChat powcoder

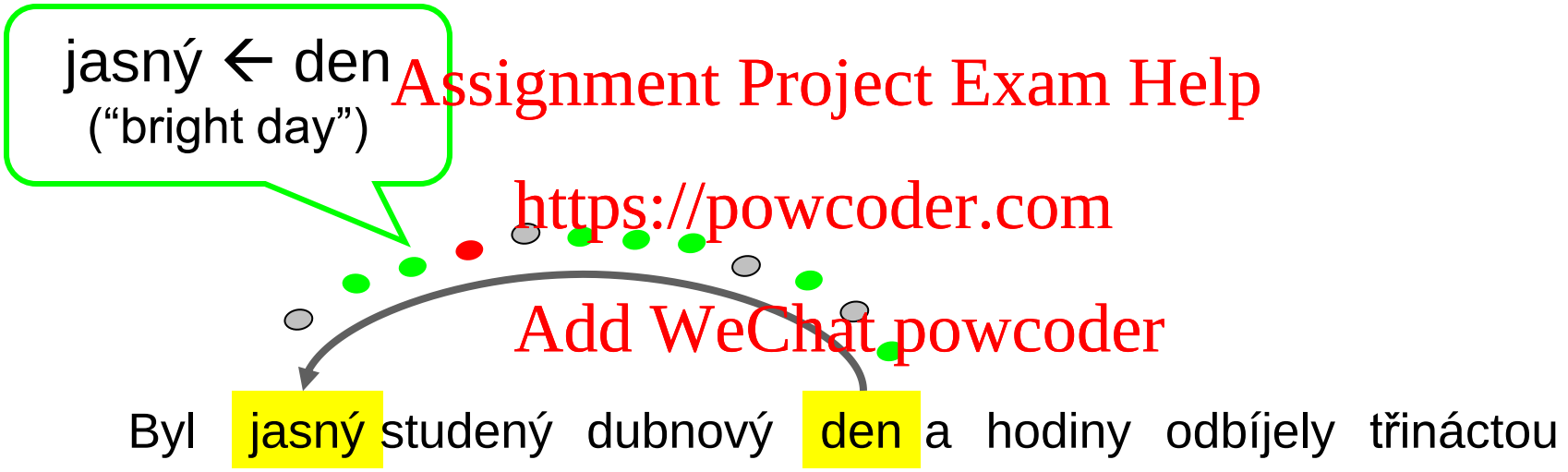


Byl jasný studený dubnový den a hodiny odbíjely třináctou

"It was a bright cold day in April and the clocks were striking thirteen"

Edge-Factored Parsers (McDonald et al. 2005)

- Is this a good edge?

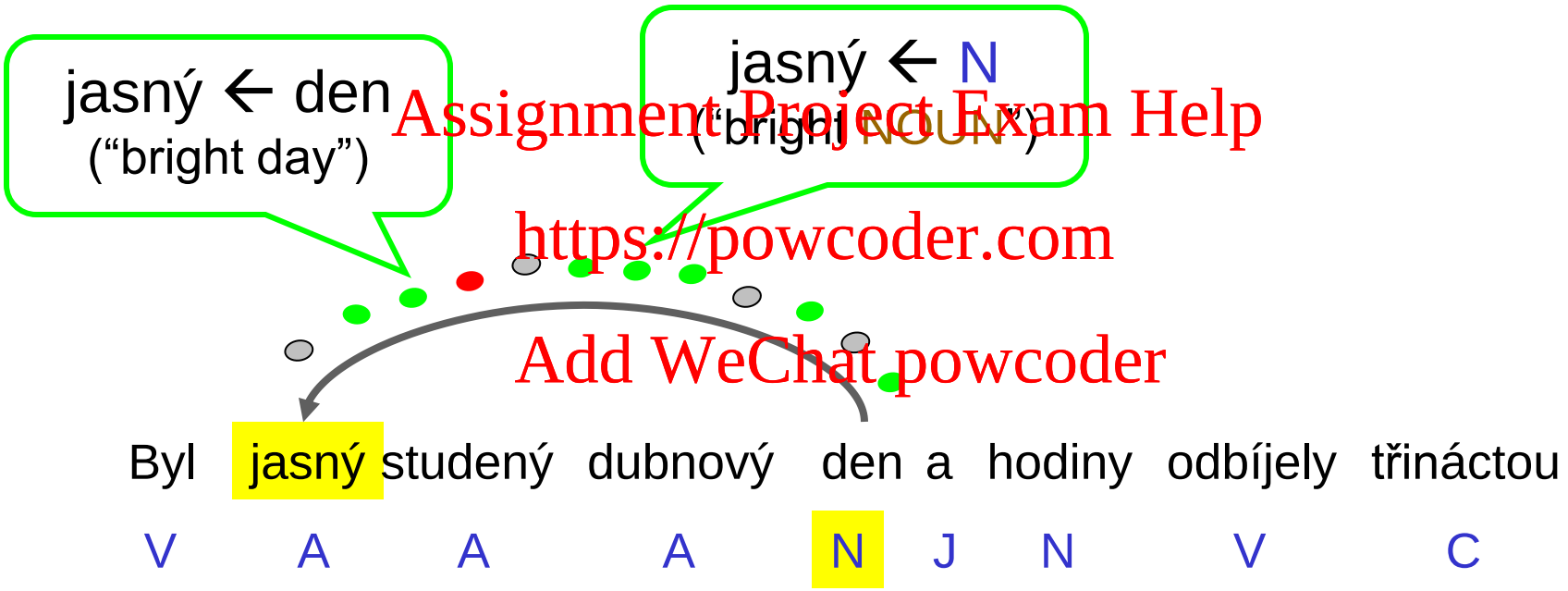


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

"It was a bright cold day in April and the clocks were striking thirteen"

Edge-Factored Parsers (McDonald et al. 2005)

- Is this a good edge?

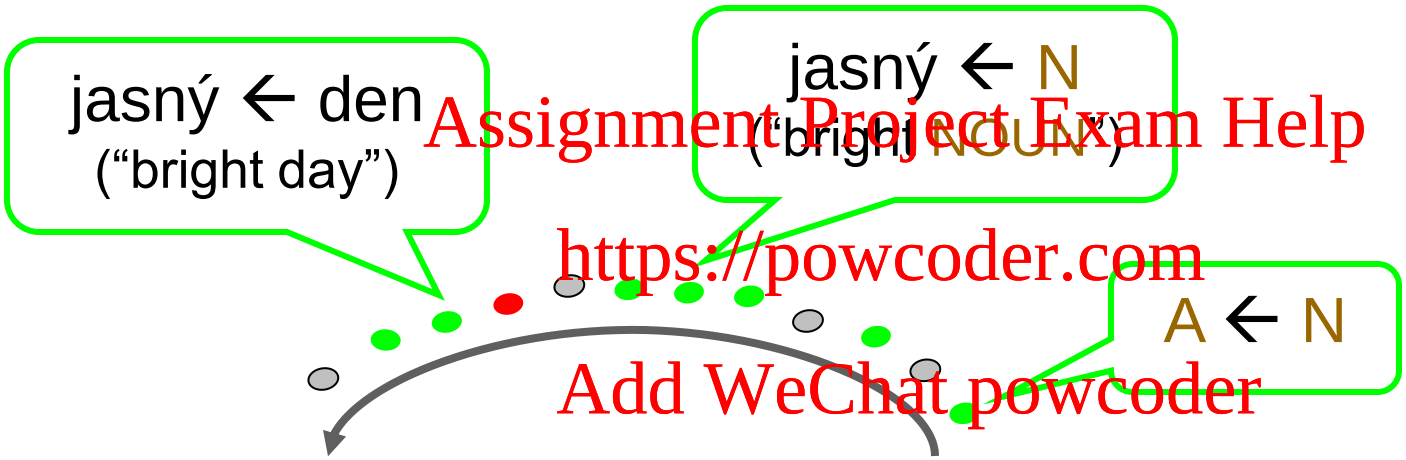


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

"It was a bright cold day in April and the clocks were striking thirteen"

Edge-Factored Parsers (McDonald et al. 2005)

- Is this a good edge?

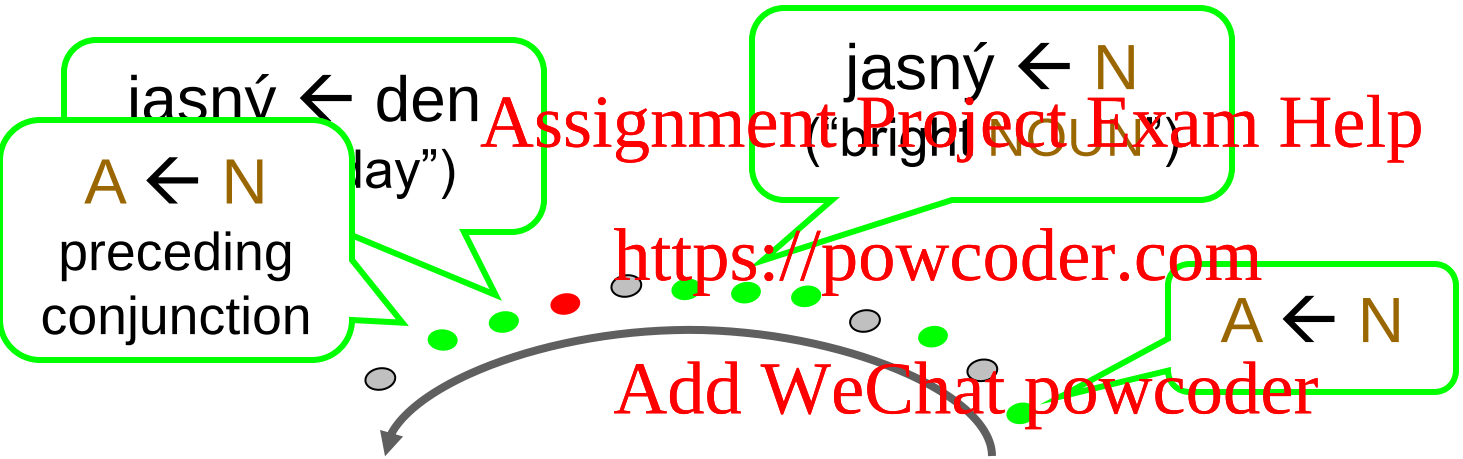


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

"It was a bright cold day in April and the clocks were striking thirteen"

Edge-Factored Parsers (McDonald et al. 2005)

- Is this a good edge?



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

“It was a bright cold day in April and the clocks were striking thirteen”

Edge-Factored Parsers (McDonald et al. 2005)

- How about this competing edge?

Assignment Project Exam Help
not as good, lots of red ...
https://powcoder.com
Add WeChat powcoder

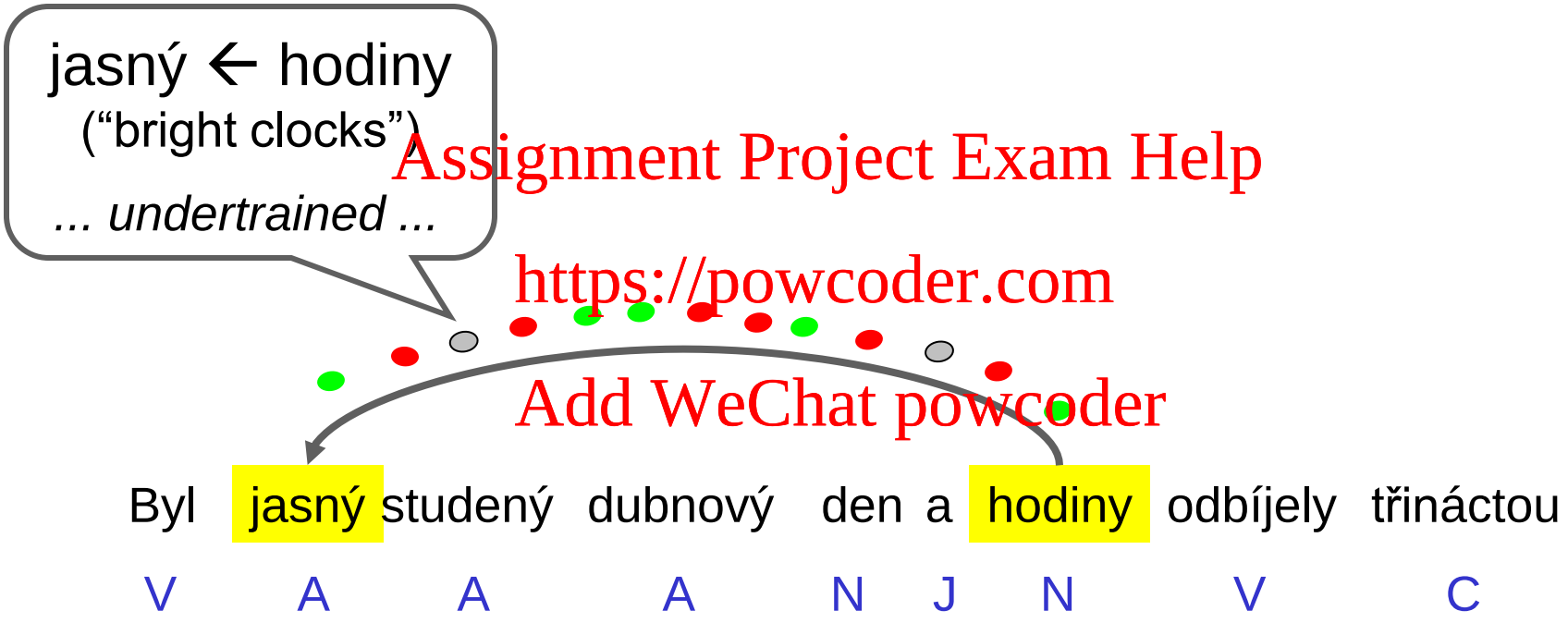
Byl jasný studený dubnový den a hodiny odbíjely třináctou

V A A A N J N V C

"It was a bright cold day in April and the clocks were striking thirteen"

Edge-Factored Parsers (McDonald et al. 2005)

- How about this competing edge?

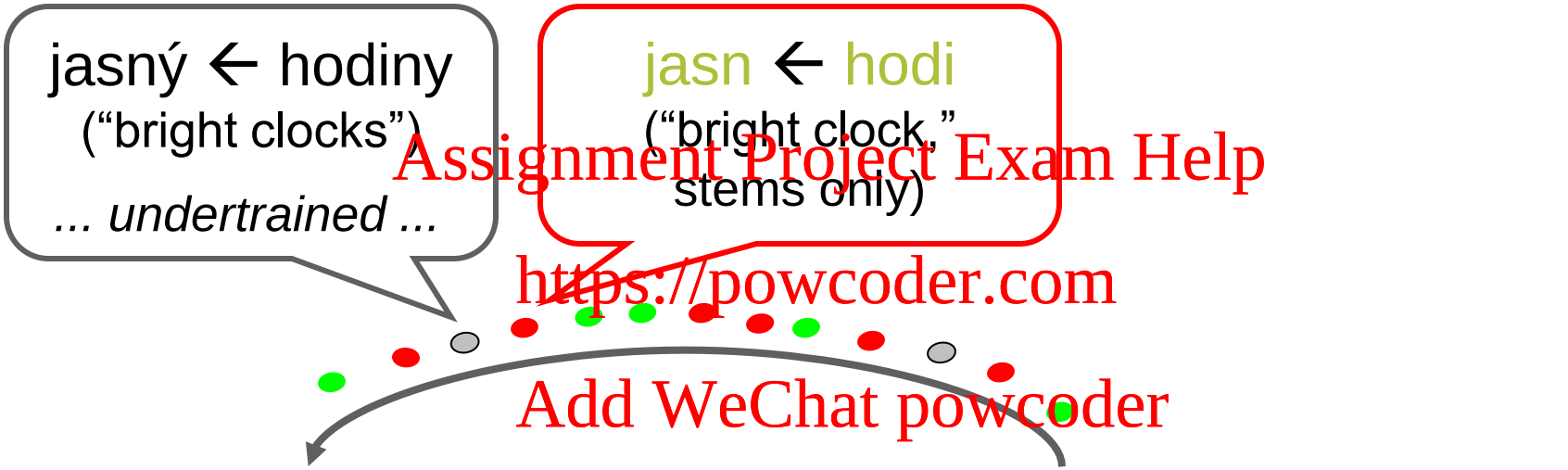


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

"It was a bright cold day in April and the clocks were striking thirteen"

Edge-Factored Parsers (McDonald et al. 2005)

- How about this competing edge?

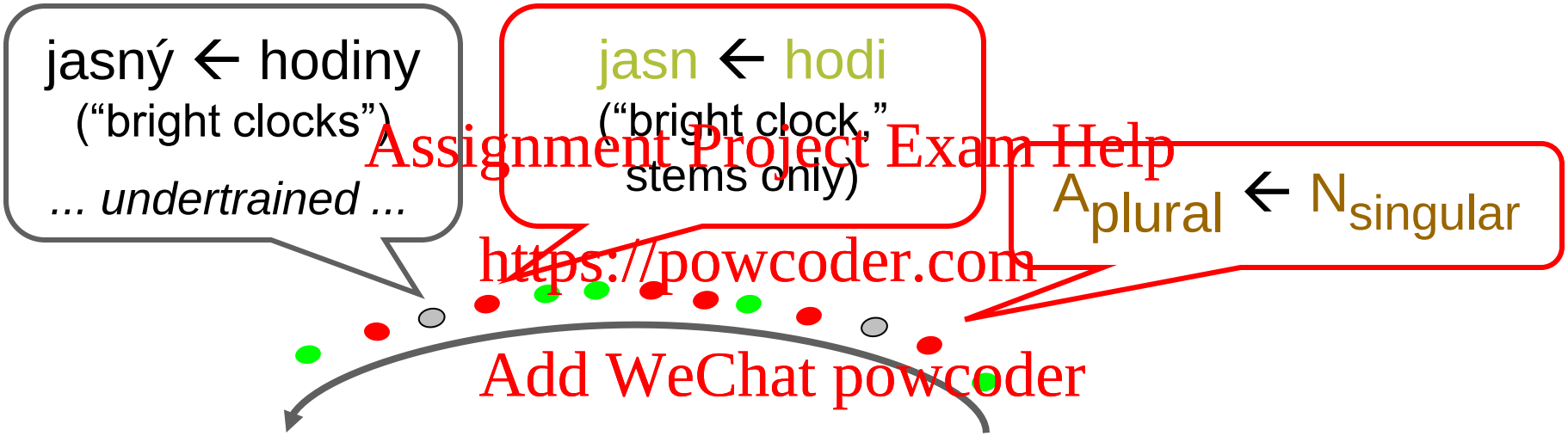


Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

"It was a bright cold day in April and the clocks were striking thirteen"

Edge-Factored Parsers (McDonald et al. 2005)

- How about this competing edge?



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Edge-Factored Parsers (McDonald et al. 2005)

- How about this competing edge?

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

"It was a bright cold day in April and the clocks were striking thirteen"

Edge-Factored Parsers (McDonald et al. 2005)

- Which edge is better?
  - “bright day” or “bright clocks”?

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Byl	jasný	studený	dubnový	den	a	hodiny	odbíjely	třináctou
V	A	A	A	N	J	N	V	C
byl	jasn	stud	dubn	den	a	hodi	odbí	třin

“It was a bright cold day in April and the clocks were striking thirteen”

Edge-Factored Parsers (McDonald et al. 2005)

our current weight vector

- Which edge is better?
- Score of an edge e = θ · **features**(e)
- Standard algos ➔ valid parse with max total score

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Byl jasný studený dubnový den a hodiny odbíjely třináctou

V A A A N J N V C

byl jasn stud dubn den a hodi odbí třin

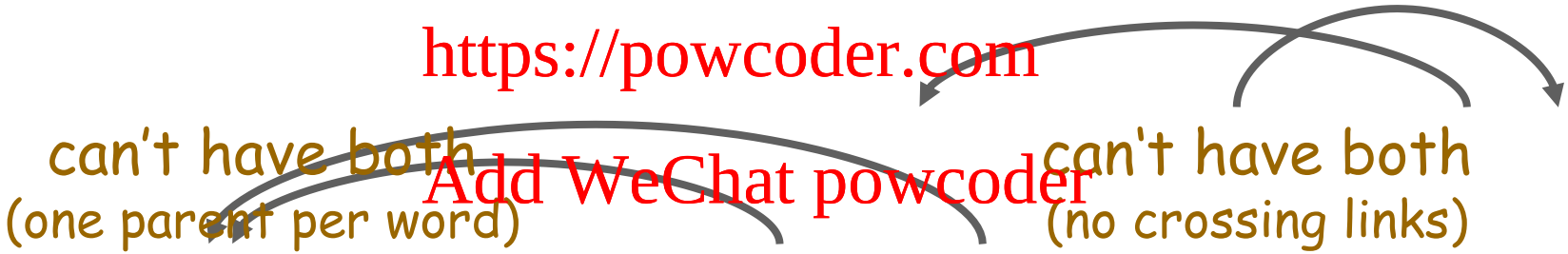
"It was a bright cold day in April and the clocks were striking thirteen"

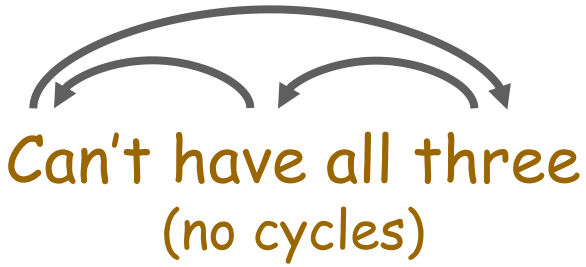
Edge-Factored Parsers (McDonald et al. 2005)

our current weight vector

- Which edge is better?
- Score of an edge e = θ · **features**(e)
- Standard algos ➔ valid parse with max total score

Assignment Project Exam Help





Thus, an edge may lose (or win) because of a consensus of **other** edges.

Finding Highest-Scoring Parse

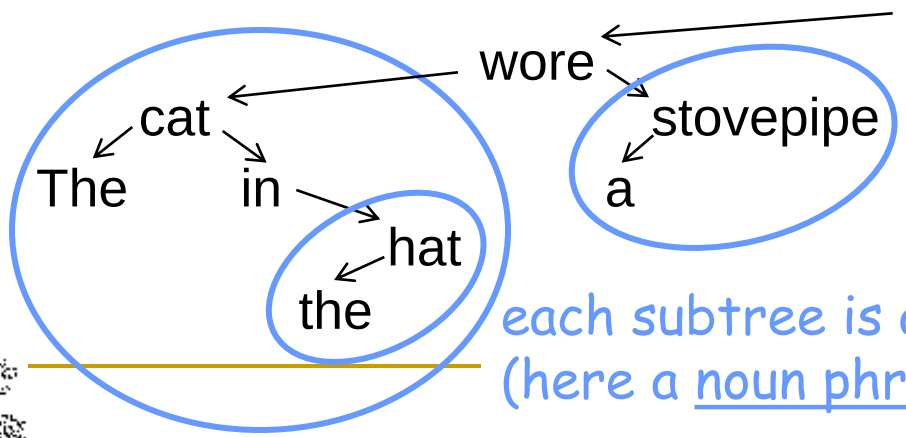
- Convert to context-free grammar (CFG)
- Then use dynamic programming

Assignment Project Exam Help

https://powcoder.com

let's vertically stretch this graph drawing

Add WeChat powcoder

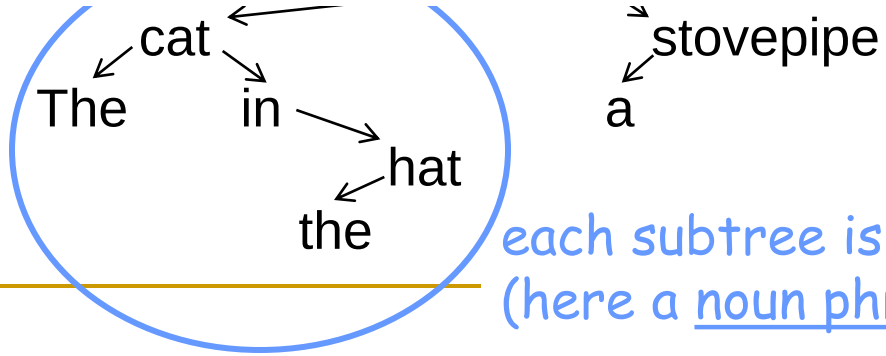


each subtree is a linguistic constituent (here a noun phrase)

Finding Highest-Scoring Parse

- Convert to context-free grammar (CFG)
- Then use dynamic programming
  - CKY algorithm for CFG parsing is $O(n^3)$
  - Unfortunately, $O(n^5)$ in this case
    - to score "cat + wore" link, not enough to know this is NP
    - must know it's rooted at "cat"
    - so expand nonterminal set by O(n): $\{NP_{the}, NP_{cat}, NP_{hat}, ...\}$
    - so CKY's "grammar constant" is no longer constant ☹

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

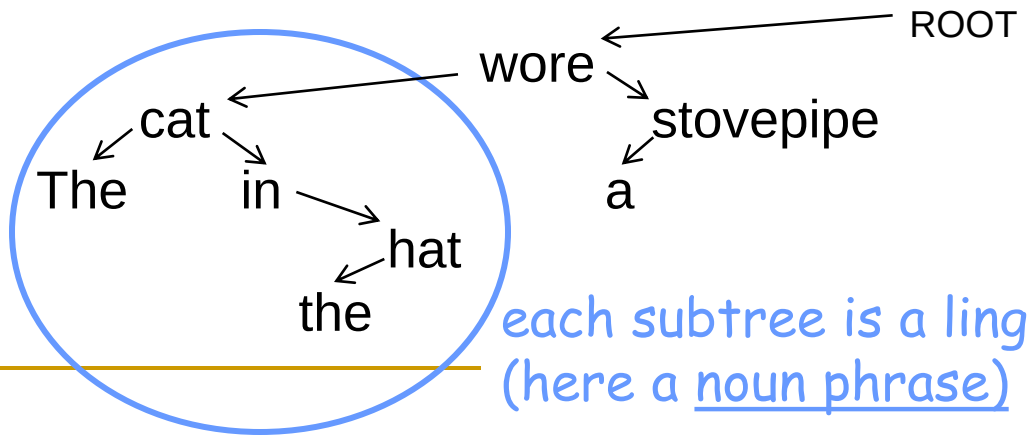


each subtree is a linguistic constituent (here a noun phrase)

Finding Highest-Scoring Parse

- Convert to context-free grammar (CFG)
- Then use dynamic programming
  - CKY algorithm for CFG parsing is $O(n^3)$
  - Unfortunately, $O(n^5)$ in this case
  - Solution: Use a different decomposition (Eisner 1996)
    - Back to $O(n^3)$

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



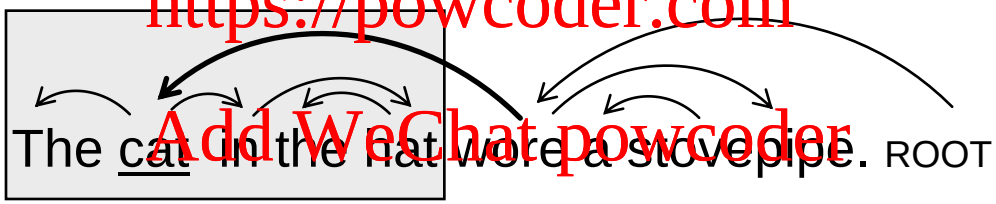
each subtree is a linguistic constituent (here a noun phrase)

Spans vs. constituents

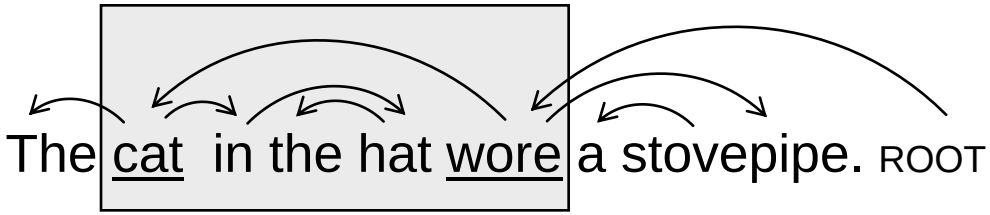
Two kinds of substring.

» **Constituent** of the tree: links to the rest only through its headword (root).

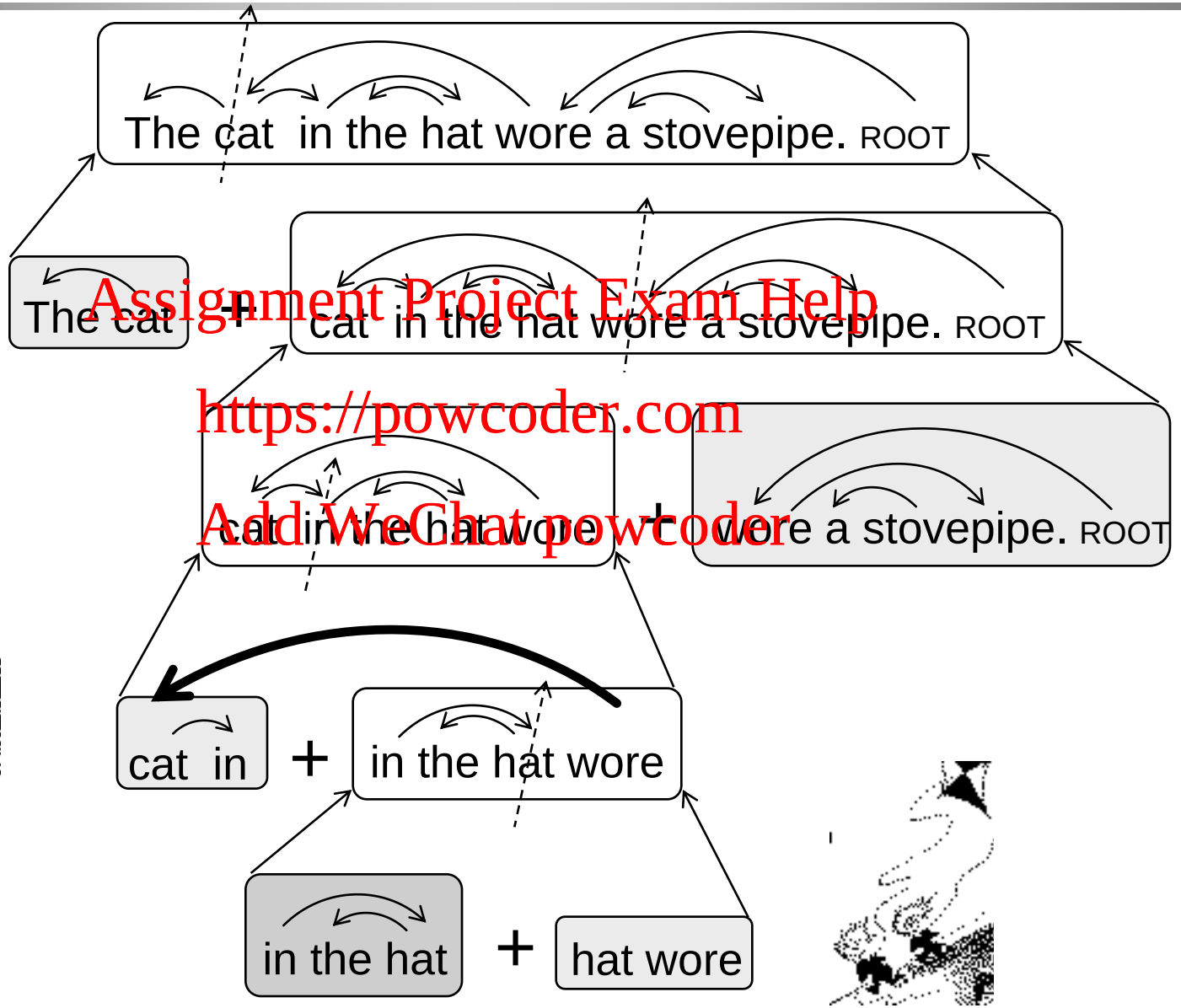
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



» **Span** of the tree: links to the rest only through its endwords.



Decomposing a tree into spans



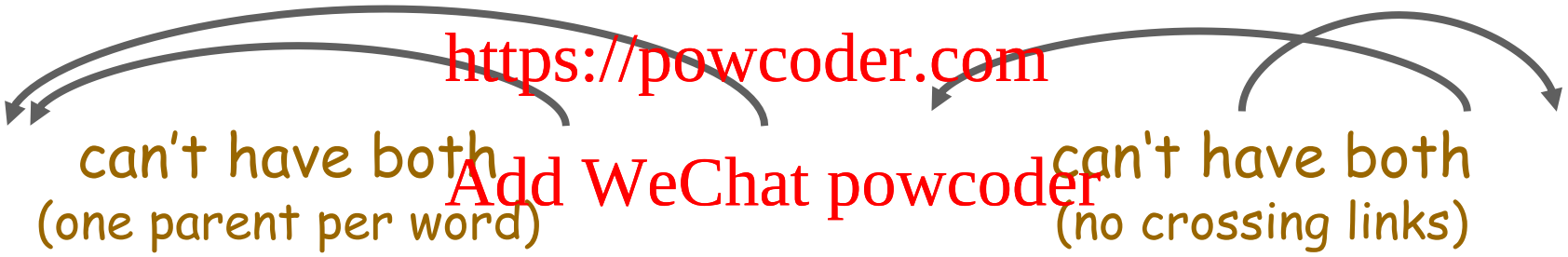
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

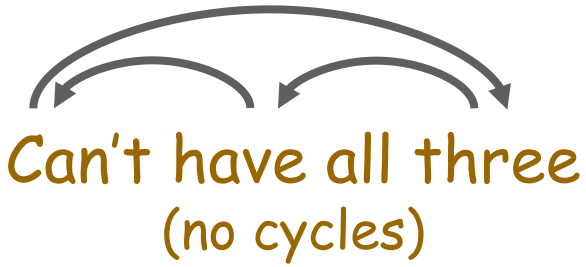
Hard Constraints on Valid Trees

our current weight vector

- Score of an edge e = $\theta \cdot$ **features**(e)
- Standard algos ➔ valid parse with max total score

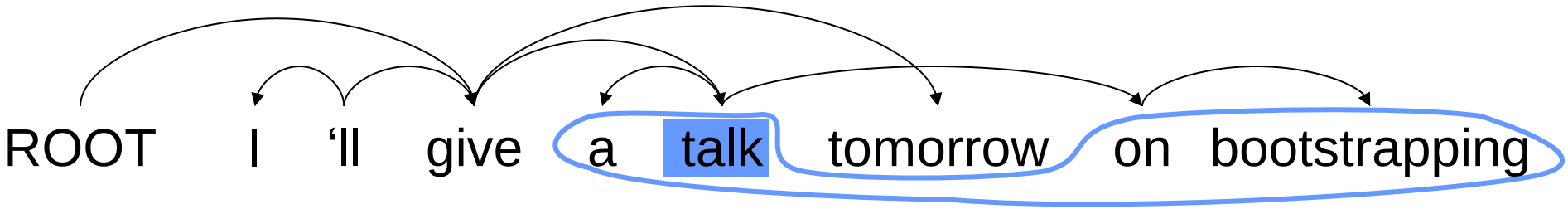
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder





Thus, an edge may lose (or win) because of a consensus of **other** edges.

Non-Projective Parses



subtree rooted at "talk"
is a discontiguous noun phrase

Assignment Project Exam Help

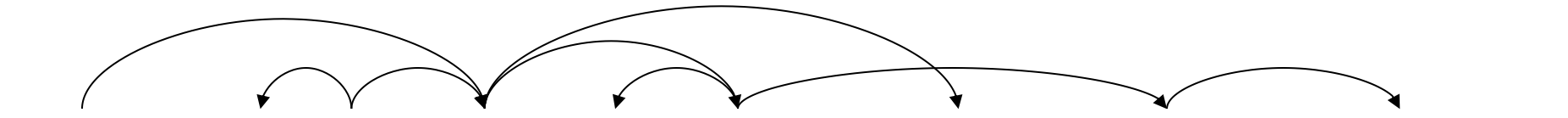
https://powcoder.com

Add WeChat powcoder

can't have both
(no crossing links)

The "projectivity" restriction.
Do we really want it?

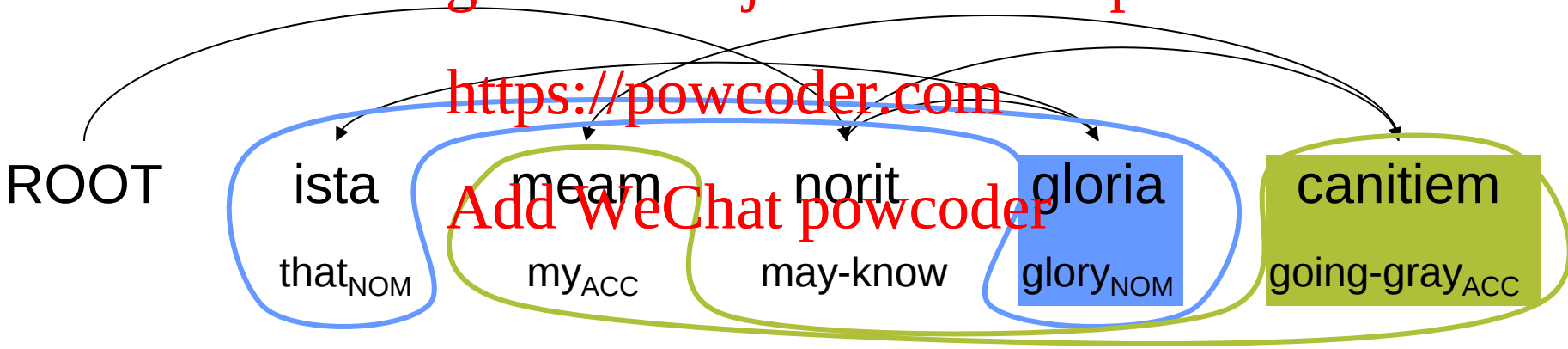
Non-Projective Parses



occasional non-projectivity in English

Assignment Project Exam Help

https://powcoder.com



Add WeChat powcoder

That glory may-know my going-gray
(i.e., it shall last till I go gray)

frequent non-projectivity in Latin, etc.

Non-Projective Parsing Algorithms

- Complexity considerations:
  - Projective (Proj)
  - Non-projective (NonP)

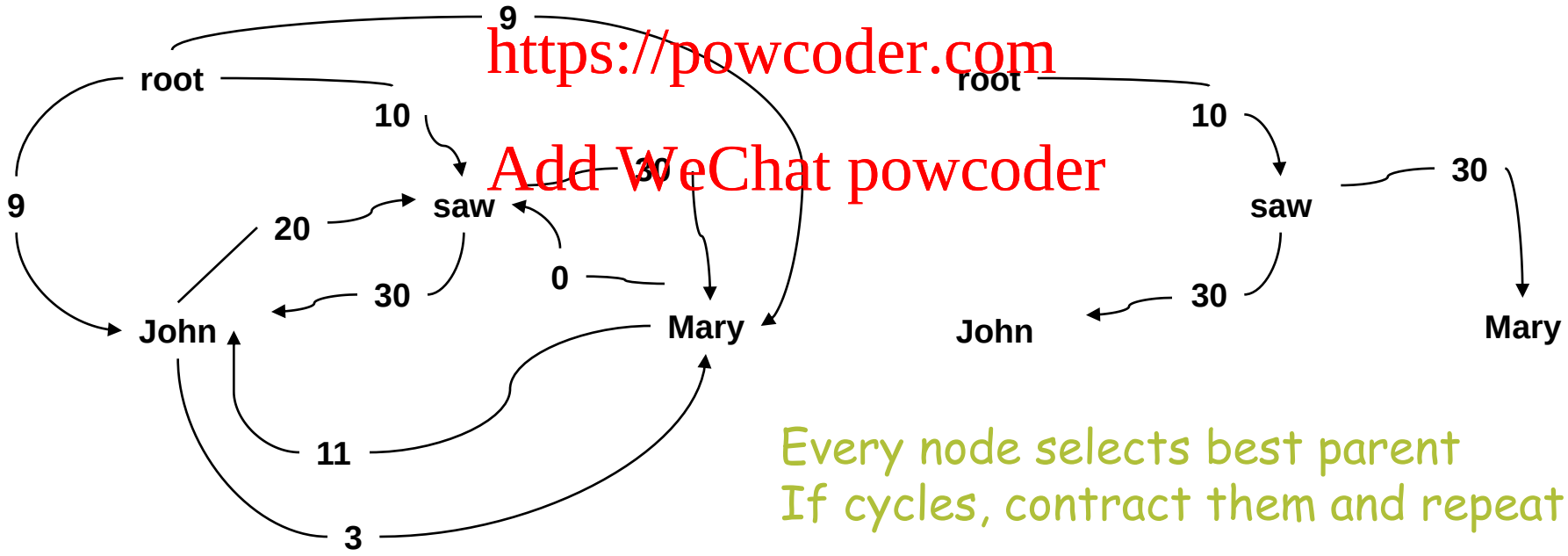
Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Problem/Algorithm	Proj	NonP
Complete grammar parsing [Gaifman 1965, Neuhaus and Bröker 1997]	P	NP hard
Deterministic parsing [Nivre 2003, Covington 2001]	$O(n)$	$O(n^2)$
First order spanning tree [McDonald et al. 2005b]	$O(n^3)$	$O(n^2)$
Nth order spanning tree ($N > 1$) [McDonald and Pereira 2006]	P	NP hard

McDonald's Approach (non-projective)

- Consider the sentence "John saw Mary" (left).
- The Chu-Liu-Edmonds algorithm finds the maximum-weight spanning tree (right) – may be non-projective.
- Can be found in time $O(n^2)$.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



Every node selects best parent
If cycles, contract them and repeat

slide thanks to Dragomir Radev

Summing over all non-projective trees

~~Finding highest-scoring non-projective tree~~

- Consider the sentence “John saw Mary” (left).
- The Chu-Liu-Edmonds algorithm finds the maximum-weight spanning tree (right) – may be non-projective.
- Can be found in time $O(n^3)$.

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

- How about total weight Z of all trees?
- Can be found in time $O(n^3)$ by matrix determinants and inverses (Smith & Smith, 2007).

slide thanks to Dragomir Radev

Graph Theory to the Rescue!

$O(n^3)$ time!

Tutte's **Matrix-Tree Theorem** (1948)

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

The **determinant** of the Kirchoff (aka Laplacian) adjacency matrix of directed graph *G* without row and column *r* is equal to the **sum of scores of all directed spanning trees** of *G* rooted at node *r*.

Exactly the *Z* we need!

Building the Kirchoff (Laplacian) Matrix

$$\begin{vmatrix} \sum_{j \neq 1} s(1,j) & -s(2,1) & \cdots & -s(n,1) \\ -s(1,2) & \sum_{j \neq 2} s(2,j) & \cdots & -s(n,2) \\ \vdots & \vdots & \ddots & \vdots \\ -s(1,n) & -s(2,n) & \cdots & \sum_{j \neq n} s(n,j) \end{vmatrix}$$

- Negate edge scores
- Sum columns (children)
- Strike root row/col.
- Take determinant

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

N.B.: This allows multiple children of root, but see Koo et al. 2007.

Why Should This Work?

Clear for 1x1 matrix; use induction

$$\begin{vmatrix} \sum_{j\neq 1} s(1,j) & -s(2,1) & \cdots & -s(n,1) \\ -s(1,2) & \sum_{j\neq 2} s(2,j) & \cdots & -s(n,2) \\ \vdots & \vdots & \ddots & \vdots \\ -s(1,n) & -s(2,n) & \cdots & \sum_{j\neq n} s(n,j) \end{vmatrix}$$

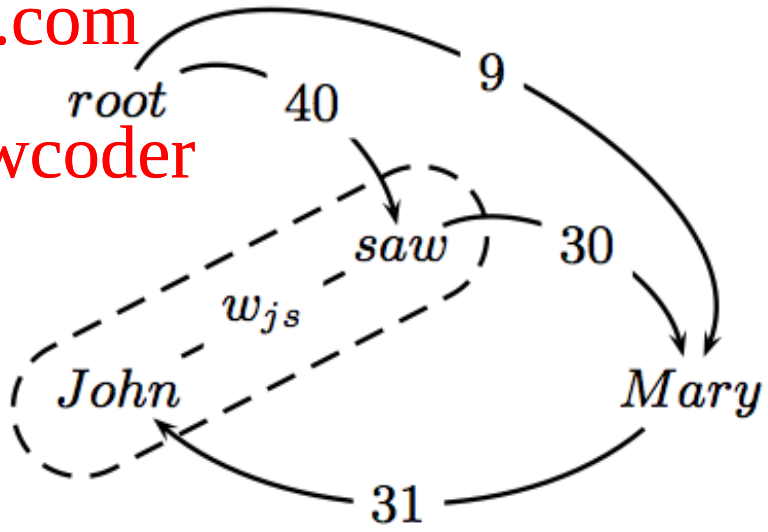
$K' \equiv K$ with contracted edge $1,2$

$K'' \equiv K(\{1,2\} \mid \{1,2\})$

$|K| = s(1,2)|K'| + |K''|$

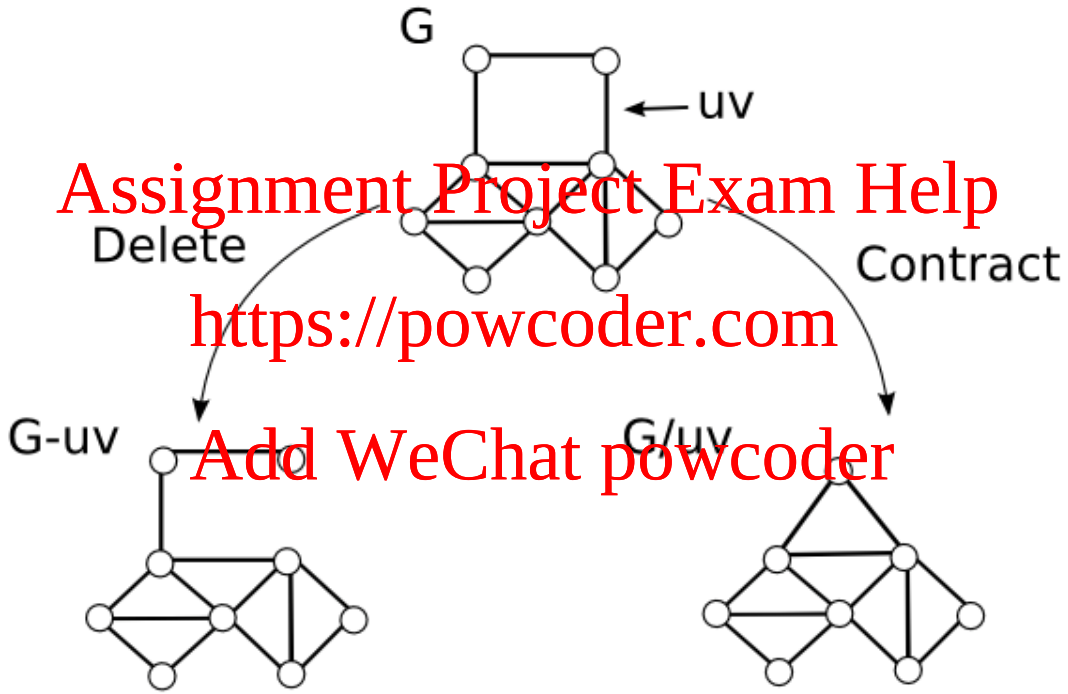
Chu-Liu-Edmonds analogy:
Every node selects best parent
If cycles, contract and recur

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



Undirected case; special root cases for directed

Graph Deletion & Contraction



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Important fact: $\kappa(G) = \kappa(G-\{e\}) + \kappa(G\backslash\{e\})$